APPLICATION OF IMAGE ANALYSIS IN EXTERNAL AND INTERNAL QUALITY ASSURANCE FOR DIAGNOSTIC CLINICAL IMMUNOHISTOCHEMISTRY

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Masters of Science
in the Department of Health Sciences
University of Saskatchewan Saskatoon

By
Heather Neufeld

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

DISCLAIMER

The formalin-fixed, paraffin-embedded cell line microarrays were exclusively created to meet the thesis and/or exhibition requirements for the degree of Master of Science at the University of Saskatchewan. Reference in this thesis/dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

The Health Sciences Graduate Program in the College of Medicine
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5
Canada

ABSTRACT

Clinical immunohistochemistry (IHC) techniques are not yet fully standardized. In this project, a standardization method was developed and tested for proficiency testing (PT) in external quality assurance (EQA) and quality control (QC) in clinical IHC laboratories. The breast cancer markers estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor 2 (HER2) were used as a model system. Digital image analysis (IA) was used in conjunction with new calibrated and standardized cell line microarrays (CLMA). CLMAs built from nine formalin-fixed paraffin-embedded (FFPE) breast cancer cell lines were used for both QC controls and PT samples, instead of traditionally used FFPE tissues, in the standardization of breast cancer IHC. IA was used for measurement of IHC results, and compared to evaluation by the traditional expert-assessment method.

Laboratory Score: Reference Score Ratio (LSRSR) was derived from Histo-Scores (HScores) determined by IA. HScores and LSRSRs were examined statistically and evaluated as histograms and boxplots to summarize and rank participant laboratory EQA results, in comparison to a reference sample or reference laboratories in two consecutive Canada-wide EQA runs.

LSRSR-derived reference ranges were highly sensitive in evaluating laboratory EQA performance in PT as well as for monitoring of controls for QC. Laboratory on-slide tissue and cell-line IHC QA controls were assessed using IA and Levey Jennings QC charts. These charts were determined to be an excellent way to observe trending in laboratory IHC staining over time, particularly when cell line controls were used. This approach also reduced the time and labor costs for PT evaluation.

Overall, cell line calibration controls were functionally equivalent or better than tissue-based controls in QC and PT mainly because of cell line biological homogeneity and sample availability. This study identified an optimal design for preparation of IHC cell line controls and PT samples for breast cancer markers. Optimal, intermediate staining cell line IHC controls were identified for all three breast cancer markers.

Using IA with LSRSR and cell line samples is recommended for standardization of IHC methodology. This approach advances QA for diagnostic IHC and when implemented will improve patient care.

ACKNOWLEDGEMENTS

I want to take this opportunity to sincerely thank the many individuals who supported and encouraged me during my research:

To begin, I am indebted to my co-supervisors Emina Torlakovic and Barry Ziola for their support, patience and guidance. My work has been immeasurably enhanced by their mentorship and encouragement.

I am also grateful to Blake Gilks, John Garratt, Dragana Pilavdzic, Shannon Klassen, and all of the participating laboratories of the Canadian Immunohistochemistry Quality Control Program, for their advice and cooperation in this project. It was a rare opportunity to work with a national quality assurance program to gain practical, applied knowledge and gather data in immunohistochemistry quality assurance.

I acknowledge Deborah Anderson, Svein Carlsen, John Decoteau and Erique Lukong for their kind donation of the breast cancer cell lines used in my research. Thank you to Bill Laverty for his advice regarding appropriate statistical analysis. Thank you, also, to the Bernard Laboratory, at the Huntsman Cancer Institute of the University of Utah, for their skilled RT-qPCR analysis.

I sincerely thank the Department of Pathology and Laboratory Medicine and the College of Health Sciences Deans at the University of Saskatchewan for their understanding and support while I worked with them in parallel to completing my research and thesis. I also acknowledge my committee, Svein Carlsen, John Krahn, Lou Qualtiere and Anurag Saxena, for their encouragement and advice.

I must also thank my husband James for his unfailing love, encouragement, and patience throughout this project. I am eternally grateful for my family's support and encouragement.

I dedicate my thesis to the memory of my mother, Yvonne Baxter, who was not able to see this project completed. She encouraged me to approach all things with compassion and perseverance. I hope she would be proud.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

AR, antigen retrieval

ASCO, American Society of Clinical Oncology

ATTC, American Tissue Type Collection

CAP, College of American Pathologists

CCD, charge-coupled device

CAP-ACP, Canadian Association of Pathologists

cIQc, Canadian Immunohistochemistry Quality Control

CLMA, cell line microarray

DAB, diaminobenzidine

D-PBS, Delbucco's phosphate buffered saline (calcium and magnesium free)

EDTA, ethylenediaminetetraacetic acid

ER, estrogen receptor

EQA, external quality assurance

FBS, fetal bovine serum

FDA, Food and Drug Administration (United States)

FFPE, formalin-fixed paraffin-embedded

FISH, fluorescent *in situ* hybridization

HER2, human epidermal growth factor receptor 2

HIER, heat induced epitope retrieval

HRP, horseradish peroxidase

HScore, Histo-Score

IA, image analysis

IHC, immunohistochemistry

LSRSR, Laboratory Score: Reference Score Ratio

JGH, Jewish General Hospital (Montreal, QC)

PR, progesterone receptor

PT, proficiency testing

QA, quality assurance

QC, quality control

RGB, red, green and blue

RT-qPCR, reverse transcription quantitative real-time polymerase chain reaction

SCH, Saskatoon City Hospital

SD, standard deviation

SHR, Saskatoon Health Region

T25, 25 cm^2

T75, 75 cm^2

TMA, tissue microarray

VGH, Vancouver General Hospital

κ = Cohen's Kappa value

1.0 LITERATURE REVIEW

Compared to other clinical laboratory methods, clinical immunohistochemistry (IHC) is not fully standardized. IHC daily quality control (QC), which is a part of laboratory quality assurance (QA), is not well developed because the required standardized positive and negative controls are not yet available. Similarly, proficiency testing (PT) is not standardized as a part of external quality assurance (EQA), despite being an essential part of inter-laboratory comparison, test calibration, and national QA standardization. The aim of QA is to ensure test results are reproducibly accurate. This is essential for appropriate diagnosis and patient care. The aim of PT is to demonstrate that clinical and diagnostic laboratories perform at the same or comparable level as reference laboratories.

The research described here represents the first attempt to develop standardized controls for use in both QC and PT IHC. Breast cancer marker testing by IHC was the focus of this research due to the significant clinical implications for patients if test results are inaccurate [1]. Test inaccuracy can result from various pre-analytical, analytical, and post-analytical factors. In diagnostic IHC, the pre-analytical component is defined as human tissue processing that occurs in the operating room and pathology departments, until the sample reaches the immunohistochemistry laboratory. The analytical component of IHC includes the antigen retrieval (AR) step and the application of antibodies and staining agents. The post-analytical component of IHC is the interpretation of results. Many expert committees are currently addressing both pre-analytical and post-analytical components in an ongoing effort to standardize these components. However, it is very difficult to truly standardize the analytical component of IHC, since methods vary from one laboratory to another and there is continuous effort to improve IHC technology and automation. Consequently, current emphasis is being placed on the standardization of the expression of IHC results, rather than standardization of methods. The standardization of IHC results begins with the standardization of the controls used for QC and the samples used in PT. These are principal components of EQA for the external monitoring of laboratory performance.

1.1 Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) in breast cancer diagnosis

1.1.1 Breast cancer

Breast cancer is presently the second leading cause of cancer-related death in women and is the most common cancer among women excluding non-melanoma skin cancers. According to the American Cancer Society, about 1.3 million women will be annually diagnosed with breast cancer worldwide and about 465,000 will die from the disease [2].

Biologically, breast cancer is an uncontrolled growth of breast cells. Usually breast cancer begins in cells of the lobules or ducts. Cancers developing from epithelial tissues are designated as carcinomas. There are several histological subtypes of breast carcinoma [3]. Less commonly, breast cancer can begin in stromal tissues, including the fatty and fibrous connective tissues of the breast. Malignant tumors of stromal tissues are designated as sarcomas. Cancer occurs as a result of mutations in the genes responsible for regulating cell growth. Normally, cells replace themselves through an orderly growth process, with healthy new cells taking over as old ones die due to apoptosis or "programmed cell death". Various mechanisms may lead to cancer development. Some are related to deregulated growth and proliferation, while others affect cell survival [4–6]. Mutations in the apoptotic pathway allow a damaged or mutated cell to keep dividing uncontrollably, forming both benign and malignant tumors. While some cancers are due to an inherited genetic abnormality, many breast cancers have no family history. Tumors may arise as a result of an inherited predisposition for cancer combined with acquired genetic changes from the aging process [2,3,5].

Over time, cancer cells can invade nearby healthy breast tissue and make their way into the lymph nodes, before moving to other parts of the body. The spreading of cancer into lymph nodes or distant organs is called metastasis. "Staging" of breast cancer refers to how far the cancer cells have spread beyond the original tumor and includes parameters related to tumor size, tumor extension into surrounding organs, and the status of metastases. Detailed description of breast cancer staging is provided in the 2009 *AJCC Cancer Staging Manual* [7].

1.1.2 Estrogen, progesterone, human epidermal growth factor and breast cancer

Estrogen and progesterone have many important biological functions and travel through the bloodstream to receptor sites in both healthy and cancerous cells. Breast-cell growth stimulation is the most important function of estrogen and progesterone. Like normal breast tissue, many breast cancers are hormone-dependent, with estrogen and progesterone binding to ER and PR, respectively, thus stimulating cancer growth. Breast cancers can be ER-positive (ER+) and PR-positive (PR+) or ER+ only. It is rare that breast cancer is PR+ only. ER and PR testing, by IHC, are used to determine if a breast cancer expresses ER and/or PR, and will therefore respond to hormonal therapy. Women whose cancers are PR+, but ER-, still have about a 10% chance of responding to hormonal therapy [2,8].

The binding of human epidermal growth factor to HER2 initiates cell proliferation and indirectly affects programmed cell death. Between 20-25% of breast cancers have increased HER2 expression, which is related to more rapid tumor growth, an increase of recurrence after surgery and a poor clinical outcome [9]. Because of this, HER2 testing is critical in determining which cancer patients are eligible for treatment with trastuzumab (herceptin), which acts by blocking the binding site of HER2 in tumors. The College of American Pathologists (CAP) recommends that HER2 status should be determined for all invasive breast cancers [10]. HER2 detection can be done by either IHC or fluorescent *in situ* hybridization (FISH). Although both methods are very good, IHC is simpler and more cost effective. FISH is generally reserved for cases scored as 2+ by IHC. Such tumors may or may not have an amplified HER2 gene and FISH is used for confirmation of over-expression [9].

1.1.3 ER, PR, and HER2 testing and breast cancer diagnosis

Breast cancer survival has improved by approximately 25% over the past two decades. This is due, in part, to advances in the understanding of breast cancer pathogenesis and therapies targeting ER, PR, and HER2. Highly specific and sensitive IHC testing of tumor samples for ER, PR and HER2 forms the basis of stratifying breast cancer patients to different treatments. However, these IHC tests have historically suffered from poor reproducibility [8,11–16].

Proper ER and PR detection is crucial. The more receptors in a tumor, the better an individual's response will be to hormone therapy. Also, the response to the treatment could depend on whether one or both receptors are present in the tumor. For a person with ER+/PR+ status, the chance of responding to hormonal therapy is approximately 70%. With ER+/PR– or ER–/PR+ status, there is approximately a 33% chance of responding. If the receptor status is low or unknown, there is about a 10% chance of responding [17]. Therefore, detection and quantitation of ER and PR needs to be accurate. There is considerable published evidence showing that measured levels of ER and PR in given samples differ from one testing center to another [7,18–26]. Until recently, some 30% of participating clinical laboratories failed to detect ER and PR in tumors that tested positive in reference laboratories [12]. Better method controls for IHC EQA and daily QC are required to improve test accuracy.

ER and PR IHC requires 1% positivity in nuclear staining of tumor cells for a breast tumor to be declared ER+ or PR+, resulting in the patient receiving hormonal therapy. The same tumor stained for HER2 requires an overall complete membranous staining intensity level of 2 out of 3 (on a scale of 0-3+) in order for the patient to be considered a valid candidate for Herceptin treatment. These readings can be variable, depending on the skill and training of the pathologist, and on the quality of the stain itself [27].

1.2 IHC

IHC has been used routinely for over three decades in pathology and cytology diagnostic laboratories. IHC is used in both Class I and Class II diagnostic testing. Class I tests aid in diagnosing disease by the pathologist, while Class II tests also provide prognostic and predictive information (e.g., about tumor type) to be applied for appropriate stratification of patients for various therapies [10]. Standardization of clinical IHC lags behind many other clinical laboratory methods. Since IHC is not fully standardized, IHC-staining quality can vary greatly between laboratories, depending on the expertise of the staff and the methods used [28,29].

IHC procedures are complex, with many steps and methodologies available to achieve the final staining result. Research methods have been adapted for diagnostic use to visualize a protein of interest through antigen-antibody interactions. Antibody-

protein binding properties already in use in immunochemical techniques like western blotting and enzyme-linked immunoassays have been applied in IHC to directly observe the presence and localization of proteins/epitopes of interest in cells and tissues. Samples can be either cytological preparations on slides, or sections of either frozen tissue or formalin-fixed paraffin-embedded (FFPE) tissue that have been sectioned with a microtome and mounted onto microscope slides. In clinical practice, the great majority of IHC tests are performed on FFPE samples. Current attempts are underway to standardize terminology and methodology used in the IHC procedure[30–35].

In clinical analysis, an automated immunostaining system is most frequently used for IHC staining. These instruments allow for high throughput, using proprietary reagents that include a labeled secondary antibody for visual detection. FFPE samples are sectioned onto a slide with a microtome before they are prepared for staining with AR, usually by heat-induced epitope retrieval (HIER). IHC can be used to visualize the presence and location of individual proteins within the tissue and cellular structures of a biopsy or tissue sample. Primary antibodies are selected by individual pathologists and laboratories for their affinity to proteins (epitopes) that are important for clinical diagnosis. A labeled secondary antibody is used to detect the binding of a primary antibody to the antigen of interest. Visual identification involves using one of a number of different chromagen detection systems. Slides are then dehydrated and mounted with a coverslip for viewing and storage [34,36].

Although expert pathologists interpret staining intensity and antigen localization for patient diagnosis according to published literature, the clinical applications of IHC, patient tissue processing, IHC protocols, and the interpretation of IHC results, are not standardized. This produces varying IHC results in different testing centers[37]. There is also no current published consensus for the techniques and sample selection for IHC EQA and internal QC because they are still being defined [37,38].

1.2.1 Pre-analytical steps – sample collection, fixation and preparation

1.2.1.1 Fixation and formalin embedding

The pre-analytical component of IHC encompasses tissue processing from collection in the operating room or morgue until the processed tissue is ready to be

stained in the IHC laboratory. It also includes the so-called “ischemic time” during surgery and postsurgical transport to the pathology laboratory, both of which may vary greatly. There can also be significant variation in the length of fixation once a pathologist has evaluated the specimen. One current recommendation is that optimally prepared breast cancer tissue samples should be fixed for a minimum of 8 hr (to a maximum of 72 hr) before being embedded in paraffin [39,40]. Other sources recommend 12 hr for minimum fixation [14,20,33].

Although there are recommendations of fixation within 1 hr of collection, there are no binding guidelines for the proper time from excision to fixation for an IHC sample [39,41,42,43,44]. Although optimal fixation periods have been recommended for breast cancer markers, there is no evidence that these recommended conditions can be widely applied to the more than 200 other clinically important IHC markers [39,41,42]. Currently, Part I of the Canadian Association of Pathologists (CAP-ACP) IHC Checklist includes a general recommendation that all tissues should fixed >12h before IHC testing [45].

IHC pre-analytical steps vary because tissue samples are collected using numerous conditions and methods, depending on the hospital and physician collecting the sample. The tissue may be from surgery, biopsy, or autopsy [32,46,47]. The period of time between collection and proper fixation can vary greatly [48]. As a result, biochemical degradation of epitopes of interest may occur even before patient samples have been received by pathology departments for proper fixation. Recent guidelines for IHC HER2 testing have started to address these pre-analytical IHC issues for the first time [33,43,45]. Also, many hospitals have begun to mandate the monitoring and recording of IHC pre-analytical parameters for breast cancer testing [10,21,22,43,45,49].

Fixatives used in pathology laboratories can also vary. The most common preservative used is 10% buffered formalin. It gives results that physicians are most familiar with [50]. Other non-formalin based fixatives are available, but their reliability and accuracy in regular clinical use is not yet fully demonstrated [51]. The formalin itself can vary in freshness and quality, and the size of the sample will affect how quickly the fixative can diffuse through it for preservation. Diffusion itself is not sufficient for proper formalin fixation because the biochemical reactions involved in fixation take approximately 12 hours to complete [39,41,48,52]. Samples preserved for IHC should be

segmented into pieces thin enough that the fixative rapidly permeates the entire sample. This prevents the damaging effects of tissue ischemia and irrecoverable changes in epitopes. Ideally, IHC samples should be resected into small (5-10 mm thick) sections and immediately placed into small histology cassettes for fixation and eventual paraffin embedding [53] . A volume of between 10:1 and 20:1 of fixative to tissue is recommended for rapid fixative permeation of tissue [26,43,54,55]. It is also recommended that the time of collection and initiation of fixation should be recorded during hospital sample processing [29,43,56,57]. This information will assist pathologists in the identification of improperly collected, suspect samples and reduce IHC error due to improper fixation

One study reported that a minimum of 8 hr of fixation is required for quality IHC staining for the detection of ER in breast cancer. However, 24 hr of fixation is the approximate optimal fixation time for most samples prior to paraffin embedding [48]. Fixation of greater than 3 days can interfere with the recovery of some epitopes by tissue pretreatment methods, leaving epitopes inaccessible for binding with primary antibodies. Many epitopes are shown to be lost after 7 days and unmeasurable after 14 days [42,54]. Other sample pre-treatments, like decalcification, can alter IHC-staining quality in an unrecoverable manner [58]. The use of decalcified tissues for clinical IHC of ER, PR and HER2 is currently not recommended [53,58]. The effects of fixation and decalcification need to be elucidated for the different IHC epitopes being tested for in each tissue type [39,41,54,58–60].

Following fixation, tissues are dehydrated with ethanol washes and embedded in paraffin with an automated tissue processor and embedding system. Paraffin embedding became a common part of histological tissue study in the 1950s, for use with fluorescent-labeled antibodies [61]. Tissue processing and immunostaining developed through the 1980s and 1990s in histology [62–64], with automated tissue processing standardizing the paraffin embedding procedure [63]. Once embedded, the tissue can be cut into thin sections (usually 3-5 μm) onto charged glass slides for histological staining. The thickness of the sections used in clinical IHC varies greatly between testing centers because there is no current standard for section thickness.

It is important to note that although epitopes appear to be relatively stable once a sample is embedded in paraffin, there is evidence that once sections are cut from a

block, protein degradation can occur. Studies have shown that if slides are not IHC stained within 24-48 hr, they should be stored at a minimum of -20°C to prevent epitope degradation [65,66]. The best method to preserve epitopes is to cover the sections with paraffin and store them in liquid nitrogen [67]. Although the proper sectioning and storage of slides prior to IHC is very important, this still is not defined or mandated [65,66].

1.2.1.2 AR methods – HIER

Formalin fixation causes cross-linking between a protein's structural groups [68]. Cross-links of arginyl and lysyl peptide bonds and methylene bonds form between proteins and this aggregation can block many antibody binding sites. This causes the reduction, or complete elimination of, the antibody binding required for IHC staining [31,68]. Cross-linking will vary with the length of preservation and the strength and quality of the formalin solution used [54,55].

IHC can also be performed on frozen samples, which are sectioned using a microtome, with the sections placed onto glass slides. Frozen samples are not fixed prior to freezing, and are often used for the preparation of samples in which formalin will destroy the epitope of interest. Frozen sections do not contain formalin cross-link artefacts, and do not require pre-treatment or HIER before staining. Unfortunately, the freezing and thawing of tissues can disrupt tissue morphology, making the localization of the IHC staining more difficult to identify and interpret. This prevents wider use of frozen tissue in diagnostic IHC. FFPE tissues are more stable than frozen tissues and are generally preferred for histological study and diagnosis [36,69] Only a small number of clinical IHC tests are currently performed on frozen tissue sections at specialized laboratories.

IHC staining of FFPE samples was variable and inconsistent until AR was discovered and pioneered in the 1990s [68,70–72]. Initial AR methods used proteolytic digestion with trypsin, proteinase K, pronase, pepsin or deoxyribonucleotidase at varying concentrations for short periods of time. Not all antigens benefit from enzymatic digestion. Extended enzyme digestion may destroy the antigens or the tissue sample [71,73]. Although pepsin and protease are still used for some antibodies in clinical

methods, the majority of AR methods combine heat and chemical buffer incubation [68,71,73–75]. Most enzyme-based AR methods are becoming obsolete.

The most common current method of AR prior to immunostaining is HIER. HIER was first demonstrated in 1993 [74,76] and allows IHC to be functional with FFPE samples in clinical and research settings [47]. Manual methods of HIER involve the heating of samples in AR solutions to high temperatures (95° - 100°C) via water baths, microwave ovens, autoclaves and pressure cookers. Microwaves and digitally calibrated pressure cookers are most commonly used in clinical HIER [47,77,78].

Depending on the tissue and the buffer, the optimal incubation of sample in the AR buffer can vary [31,77]. Preferred solutions include citrate buffers of pH 6, 9 and 10, and 1 mM ethylenediaminetetraacetic acid (EDTA) solutions of pH 9-10 [68]. There is still debate as to the optimal HIER buffer for individual proteins [68,77]. Research has shown that the optimal buffer and buffer pH used for HIER varies from antibody to antibody. Therefore, HIER should be optimized for each antibody, in each laboratory, prior to clinical use [71,72].

The standardization of AR remains a major issue in reducing IHC-stain variability[79]. Incorrect or irregular AR can affect the quality of IHC staining, so that weakly positive samples are missed. There is still a need to define optimal sample preservation and AR methods for IHC [32,80]. In an attempt to streamline and optimize AR for clinical IHC, a number of automated AR instruments have become commercially available in the last five years. These instruments contain a controlled heat source, buffer management and humidity control. The implementation of this new technology is slow, due to high cost and lengthy validation time [36,80].

1.2.2 Methodology and stain linearity: analytical and post-analytical IHC

The analytical component of IHC encompasses the protocol used for IHC staining, which has a greater potential for standardization than the pre-analytical component of IHC. The analytical component cannot be currently standardized because new methods and new reagents continue to be developed daily. Any attempt to standardize analytical components would also lead to undesirable outcomes, because it would prevent the development of better, new antibodies, reagents, and methodologies for IHC detection.

Therefore, current efforts are best focused on both the standardization of IHC controls and developing the means to objectively monitor IHC performance and standardize IHC results with these controls.

The post-analytical component includes the interpretation of results by pathologists or by image analysis (IA). Post-analytical standardization is both expensive and complex, but is achievable through national and international consensus and proper education [7]. Examples of progress in this area include the recently published breast marker IHC interpretation/scoring guidelines from the American Society of Clinical Oncology (ASCO) and the CAP [7,20]. Current guidelines recommend using a Histo-Score (HScore), which incorporates both the percentage of positive cells and the intensity of staining [81]. HScore is discussed in Section 2.6.3.1.

1.2.2.1 Antibody use, selection and IHC method optimization

Purification of antibodies for research in the 1920s allowed further study of their structure and binding qualities [82,83]. By 1967, there were published methods for the preparation and enzyme labeling of antibodies for use in antigen localization [84]. In 1970, Sternberger *et al* [85] published a paper specifically on the use and fluorescent labeling of antibodies that initiated the field of immunohistopathology. Early methods for fluorescent antibody labeling in histopathology were published in 1978 [86]. Peroxidase and avidin-biotin labeling methods were developed to produce stable, long-lasting antibody labels. These non-fluorescently labeled antibodies are currently more commonly used with FFPE tissues in most applications [36,70].

Antibodies for human proteins/epitopes are primarily raised in mice, rabbits and goats. Initially, only polyclonal antibodies were made. Polyclonal antibodies recognize multiple epitopes on a single molecule. While this may increase staining, polyclonal antibodies are more likely to react non-specifically. The development of purified monoclonal antibodies increased staining accuracy in IHC and other immunoassay techniques due to their purity and specificity. Monoclonal antibodies are a homogeneous population of immunoglobulin directed against a single epitope, and far more specific for use in IHC staining [36,87].

The use of monoclonal antibodies, combined with a blocking step, has removed most background staining from IHC and increased its specificity [87]. However, many excellent quality polyclonal antibodies continue to be used in both research and clinical applications. Different antibodies produced to the same protein may have different binding efficiencies because they bind to different parts of the antigen of interest. The American Food and Drug Association (FDA) have approved a number of monoclonal antibodies for clinical use; however, the use of specific antibodies, or antibody clones, is not mandated.

IHC-stain quality can vary with the antibody used [36,37,47,49,72,87–89]. Different antibodies can give very different staining results depending on the epitope they detect [90,91]. The dilution of antibody used in IHC varies depending on tissue fixation, AR methods, and reagents used [91]. Therefore, it is important, in routine practice, to know the optimal IHC conditions for each tissue type and antibody [46,92]. Currently, each laboratory may choose their own antibody and optimize conditions for it, leading to a wide variation in stain quality between laboratories [32,93].

1.2.2.2 Analytical parameters of IHC automation and IHC detection systems

A large number of detection methods and automated instruments are available for clinical IHC [47]. Each manufacturer sells reagents they recommend for use with their instruments and detection systems. Each laboratory must determine their own optimal protocols for each epitope they measure with IHC. These protocols vary because each laboratory has a different knowledge of what these optimal IHC protocols should be.

Antibody binding in IHC is visualized via a labeling compound. The chromagen can be directly linked to the primary antibody or to a secondary antibody that will bind to the primary antibody [36,47,72]. The chromagen can also be linked to a polymer molecule which is bound to a secondary antibody [36]. Because fluorescent labeling is temporary, clinical IHC usually involves a stable chromagen or particulate label. Common chromagen systems include diaminobenzidine (DAB, a brown color), 3-amino-9-ethyl-carbazole (a red or brown color, depending on processing), Hanker Yates reagent, fast blue, fast red and nitroblue tetrazoleum [36,72]. Using multiple primary antibodies together with differently labeled secondary antibodies allows for dual staining on a single slide

[36,94]. IHC positive staining can be better identified when unstained tissue is counterstained with hematoxyllin. This positive staining is localized in different cell compartments (nuclear, cytoplasmic, membranous, extracellular, or any combination of these), depending on the biological distribution of the epitope of interest.

Horseradish peroxidase (HRP)-DAB staining is the most commonly used IHC detection system, with the peroxidase bound to the antibody or polymer [36]. Osmification of DAB reagent added to the slide produces an intense brown color that is alcohol resistant and resistant to fading. Because DAB is alcohol resistant, it can be used with a wide variety of FFPE samples. DAB is the most frequently used chromagen because it has a high signal-to-noise ratio and produces a specific morphological visualization[47].

Efficient, commercial protocols and chromagens have been paired with automated staining machines for clinical use by Ventana and Dako [78,87,95]. These instruments use batch-controlled reagents, with specially designed secondary antibodies. These special secondary antibodies are bound to a polymer which has a large number of HRP molecules attached to it [36]. This antibody:polymer construct provides a large number of HRP molecules for a chromagen to react with, thus increasing IHC-stain intensity. This Automation and reagent consistency improves test reproducibility [36,79].

In its early stages, IHC lacked sensitivity and specificity, largely due to problems with reproducibility [87,96]. As techniques have become more standardized, clinical utility of IHC has increased [47,87]. Greater attention is currently being given to antibody selection and optimization of the various steps in the IHC-stain process [26,72]. Many parameters must be considered during this complicated optimization process [46,47]. A suggested definition of these optimization parameters was recently provided by the Canadian Association of Pathologists National Standards Committee for IHC [45].

1.2.2.3 Non-linearity of IHC

Due to the multilayered process of IHC staining, the final staining intensity has a linear relationship with the concentration of the stained epitopes only up to a certain staining intensity. This linear range is clearly visible in other immunoassay applications such as enzyme-linked immunoabsorbant assays[97]. In practical application (unless mandated otherwise), IHC intensity is calibrated so that staining is as intense as

possible without introducing false-positive or background staining. When IHC linearity is considered, one must consider the effects of detection systems, which are based on signal amplification. Generally, current clinically applied IHC methods are intrinsically non-linear. However, there are some special conditions in which linearity may apply. In samples with low to moderate expression levels of the epitope/analyte of interest, IHC staining intensity appears to increase in a linear fashion within a certain range, with the overall staining appearing to be logarithmic in relation to epitope concentration and antibody dilution[98]. Similarly, higher antibody dilutions, which will result in decreased test sensitivity, may also produce a linear stain result [98]. While IHC allows effective measurement of the presence of an antigen, it is difficult to use IHC in a strictly quantitative or qualitative manner when interpreted manually, due to the lack of standardization and published guidelines [28]. Intensely staining tissues allow for simple identification of positive results, but these tissues have less measurable stain variation between samples. This is due to the inability of the human eye to discern small stain intensity differences at higher cumulative stain intensities. The stain signal produced may be oversaturated for lower levels of the epitope of interest. Although higher dilutions of primary antibody may produce better linearity, clinical applications of such dilutions are not recommended, since many low-expressing tissues would be completely negative with this approach. The application of IA to IHC must take this information into account when used to monitor and calibrate IHC methods.

Because IHC allows the observation of protein expression within the cell and tissue itself, it yields a greater understanding of the disease process. To use IHC in a quantitative manner requires that current pre-analytical, analytical and post-analytical problems are addressed. Only then will IHC approach the level of accuracy and precision of other biochemical and immunoassays currently used for patient diagnoses [33,58,72]. The quantitative use of IHC on FFPE tissues is now possible because of largely standardized and optimized HIER methods, although the lack of stain linearity remains an issue [99,100,101] .

1.2.2.4 Class II IHC tests

Class II IHC tests are IHC tests from which results are interpreted by pathologists

for use by physicians who are treating patients. Physicians use Class II IHC test results to stratify patients for appropriate therapies. Class II tests are either predictive of a patient's response to a particular drug, or prognostic by giving information about a patient's outcome in particular clinical setting [33,58,72,102]. Unlike IHC, most other Class II clinical laboratory results are liquid-based with optimal control systems.

1.3 PT and daily QC

The quality of clinical laboratory service is monitored with a combination of in-laboratory QA, standard operating procedures and EQA programs. The principal goal of EQA is to evaluate laboratory results and observe their agreement to a reference source, reference laboratory or designated reference result. PT is an assessment of analytical performance in comparison to peer performance or a reference system[103]. Some EQA programs provide PT and enable participants to validate their IHC protocols or monitor their calibration. There is no current standardization of PT and the design of PT methods varies widely between the different PT programs [10,33,49,59].

Daily QC includes the use of calibration standards and acceptable ranges, which are available for routinely measured analytes for most laboratory tests. Standards are reflected in the controls designed for daily use. Unfortunately, both calibration standards and acceptable clinical ranges are not precisely defined for clinical IHC tests. It is also not possible to define IHC standards from currently used, positive IHC controls[33]. This is in stark contrast to other optimized clinical tests (like glucose measurement), which are supported by standardized controls and standardized interpretation of results.

Ideally, controls monitor instrument/method performance and aid in the interpretation of results. Good controls are stable, quantifiable, universally available, and ideally inexhaustible and affordable. Standards and controls should as closely resemble the sample being measured as possible [103]. Consequently, the majority of current clinical IHC controls are prepared from previously tested patient samples. In order to measure method stain quality and consistency, tissue sections are selected that are both positive (to confirm staining) and negative (to rule out false/nonspecific staining) [37,103,104].

1.3.1 PT

PT programs periodically send specimens (which are primarily human tissue) to a

group of participating laboratories for analysis. PT programs then compare the participant results with a designated “gold standard/reference result”. The choice of the reference result is up to the discretion of the EQA program. Results are presented in a format which compares the results of each participating laboratory with the results of their peers[105]. In IHC EQA, all participating laboratories receive an unstained slide of test sample(s). Each laboratory tests the samples and is usually required to provide an interpretation of their own results. The stained slides are typically returned to the EQA program for assessment and final scoring. Most EQA programs issue a performance report for each participating laboratory. This report may contain a comparison of the participating laboratory to other participating laboratories and a reference laboratory or result[104].

PT is very valuable and informative, but it is only a part of a total QA program. Total QA is the sum of all measures in an entire system to ensure quality testing. PT compares participants to the most relevant and accurate instrument/reagent combinations to assess performance [104,105]. However, this is not sufficient for QA, since laboratory QA includes all measurements that ensure quality laboratory testing including test validation, staff training and competency assessment, standardization of operating procedures, and equipment maintenance. Good QA will address method performance and provide training and assistance for improvement [106]. Unacceptable PT results need to be further investigated to identify whether they reflect real errors in clinical testing. If errors are found, their root cause needs to be documented, so that corrective action can be taken [104,107]. In clinical EQA, the quality of the control used is essential in evaluating laboratory methodology and instrumentation. It has been argued that the closer the EQA sample for PT is to a “normal” patient, the more useful it is [30,104].

EQA employs PT to evaluate clinical laboratories. EQA has benefits that cannot be achieved by in-laboratory quality programs alone. EQA allows for "comparison of performance and results, serves as an early warning system for problems, identifies systemic kit problems, provides objective evidence of laboratory quality, indicates where to direct improvement methods and identifies training needs”[108]. EQA programs are particularly effective when they combine the evaluation of PT results with a supportive consultation when PT results are not acceptable. Such programs positively affect the

quality of care available in hospitals by improving the quality diagnostic laboratory testing [107].

It is clear that the lack of consistent, reproducible control material for IHC hinders the optimization of current daily QA and EQA. Both human tumour tissues and cell lines can be used for this purpose. However, further study is required to identify what optimal samples are for use in PT.

1.3.2 Levey Jennings plots in QA

Walter Shewhart proposed QC charts in 1931[109] as a way to monitor quality in manufacturing operations [110]. His idea[111] was modified for use as a clinical laboratory QA system by Levey and Jennings in 1950[112]. Westgard further optimized the use of Levey-Jennings charts with the creation of his "Westgard Rules" for QC charts [113,114]. His rules dictate statistical guidelines for when controls and samples in an assay must be rejected and repeated when using these QC charts [115–117]. With the aid of these rules, clinical laboratories use QC charts to monitor both accuracy and precision in their laboratory methods. This system is usually applied to liquid-based tests and is not currently applied to cell-based assays like immunohistochemistry or flow cytometry.

To create and maintain a Levey-Jennings QC Chart, a minimum of 20 readings of the same standard or control should be consecutively measured on the same instrument using the same method and reagents. The mean value of these measurements, as well as the standard deviation (SD) from the mean is then calculated. The mean, ± 2 SD and ± 3 SD are then plotted parallel to the X-axis on the chart. New control readings are recorded using this chart as a reference. New recorded values should be within 2 SD from the mean and rejected when they are outside of ± 3 SD from the mean. Out of range controls and samples must be repeated. The more precise the test method, the smaller the ± 3 SD ranges will be[112]. These charts are primarily used for internal QA and trend monitoring. QC charts are especially useful when applied during the set-up of new laboratory methods to monitor and confirm method accuracy and precision [113,115,118,119]. It is possible that similar QC charts could be applied for laboratory-to-laboratory comparison in EQA programs, or for monitoring

trends from one PT run to another with the same analyte. None of the current EQA programs or in-laboratory QA methods use QC charts for clinical IHC [45,49,102]

1.3.3 Standardization of QC in clinical IHC

Class I and Class II IHC tests, due to how their results are used, require different approaches to QA. All IHC tests are in need of standardized control materials. Such controls could be used in quantitative IHC testing for Class II markers, including ER, PR and HER2 tests. Unfortunately, these standardized, quantitative positive and negative controls are not yet available for clinical IHC. Peptide IHC controls may be suitable, but not ideal, for this purpose [101,120–122].

For IHC, it is currently recommended that tissues of known positive and negative expression are sectioned onto the slide containing clinical samples for direct comparison [123]. Variability between each unique control specimen and within each specimen block remains a problem[10]. Each tissue control block is unique, with varying levels of antigen expression throughout the depth of each block due to biological variation. Biological variation is usually larger in tumor samples than in benign, histologically normal tissues. This prevents the use of current IHC controls as quantitative controls [10,99,100]. While attempts to use tissue microarrays (TMAs) containing multiple tumor and normal tissues are useful [124], the number and types of relevant samples required to accurately test for IHC performance characteristics for ER, PR and HER2 are still not known [10,33,40,125,126].

An IHC positive control or “reagent control” shows that the system/methods used for the test work as expected. A calibration control is a standard of known value that allows the quantification of a set of known standards. Test results are then compared to these standards and quantified via this comparison [127,128]. Very few calibration controls are currently available for clinical IHC. These include commercially available cell line controls for HER2 [129] and a small number of peptide controls [101,130]. The purpose of an IHC negative control is to display how the experiment works in the absence of the measured epitope and provide a baseline value for comparison to ensure experimental consistency [127,128]. Therefore, it is not possible to standardize a method without

standardized positive controls or references; standardization of a protocol is meaningless without control standardization [117,131–133].

Manufactured peptide control slides were recently developed in the United States and were shown to be stable, reproducible, and sensitive in monitoring IHC analytical components[120,130]. Purified peptide controls are suitable to monitor the performance of automated staining instruments to ensure that the equipment and reagents are performing at the same level of calibration from one run to another[101,121]. However, this purified protein is not an accurate reflection of naturally occurring protein expression in preserved tissue samples because these proteins have not undergone tissue sample processing. Peptide controls were used successfully in 2006 in a CAP PT run for HER2 [130], but have never been used clinically. There is still debate regarding whether peptide controls can be used as clinical IHC calibration controls and how accurately they will monitor system performance [120]. In addition to this, peptide controls are very expensive and their routine clinical use is unlikely due to their price.

1.3.3.1 Canadian IHC QA

Health is under provincial jurisdiction in Canada. As a result, histology laboratories are not accredited in all provinces and there are no national standards for IHC QA/QC and no national mandated EQA IHC program [10]. The Ontario Quality Management Program for Laboratory Services carries out provincial testing and monitoring in many laboratory disciplines[134]. Unfortunately their ER, PR and HER2 IHC PT methods are only applied on a provincial level [10,19,135].

The Canadian IHC Quality Control (cIQc) Program was established in 2009 to provide a QC/QA program for IHC testing in Canada. The cIQc Program is an academically initiated, anonymous, voluntary EQA program for Canadian diagnostic IHC laboratories. Test slides are cut from TMAs created by the cIQc Program from known tissues. Stained slides are scored (self-assessed) and these results are reported through an online TMA-Scoring program for an instant comparison to a reference laboratory result. The slides are also returned to the cIQc Program along with the protocols used for staining. Laboratory performance is evaluated by a group of expert cIQc assessors[102]. Both self-assessment and expert-assessment results are posted for

analysis, together with the different protocols used. Code numbers are used for laboratory identification since participation is anonymous. Laboratories with poor results can consult with cIQc panel experts to optimize their IHC methods. Additional cIQc slides are then made available to laboratories to test modified protocols.

The cIQc Program is a provider of PT, but it has no regulatory functions. One of the largest contributions of the cIQc Program is that de-identified results of participating and reference laboratories are made available as an open-access service on the cIQc website to provide an open, independent forum for discussion and education, and an evidence-based standardization of IHC testing [10,33].

1.4 IA

The use of digital IA to interpret IHC staining is a relatively new approach to dealing with assessment inconsistency. IA is the application of computer algorithms to measure desired aspects of a white balanced digital image[136–138]. When applied to IHC, IA can reduce human error in IHC interpretation [32,139]. Recently, digital IA has been compared to expert assessment of IHC staining as a diagnostic tool in pathology and in QA. IA evaluation of IHC of ER, PR and HER2 is performed comparatively to pathologist scoring of clinical samples in these studies [139–141]. IA has recently been FDA-approved for ER, PR and HER2 IHC interpretation under specific conditions [139,142]. IA for IHC interpretation is best applied to an area of interest selected by an expert pathologist, who will select only tumor tissue to be interpreted by IA. Uniform areas with abundant tumor are easier to evaluate with IA than tissues with a combination of benign and malignant tissue or scant tumor tissue. IA erroneously analyzes the both normal and tumor cells together when the area of analysis is not preselected. For these reasons, unsupervised IA of IHC results is not clinically acceptable.

1.4.1 Development

By 1850, microscopes were high enough in quality for reliable observation of cellular structures. In 1858, the first textbook of histopathology was published. Formalin fixation of tissues for extended study became a more common practice, and is now the standard fixation method used. Histology and IHC expanded the use of microscopy in the medical field[35]. The application of IHC in personalized medicine is known as

“morphoproteomics” [143–145].

As film photography advanced, adaptors were developed so that cameras could capture microscopic images. The first digital microscopy system was described in 1997 at the University of Maryland and the Pathology Department of Johns Hopkins Hospital [146]. Image quality increased to the point that a digital image is now equal in quality to a microscopic image viewed with the human eye. Digital image capture was first clinically applied to radiology images and has become commonplace in electron microscopy and pathology microscopy[35,147].

Digital pathology involves the scanning of entire histopathology slides as digital images that are interpretable on a computer screen with data management software. The IA system can be automated and combined with internet accessibility and annotations for education and long distance, expert consultation. Modern pathology standards of optical microscopy have expanded to include digital microscopy. There are currently some 30 commercially available imaging devices varying in image magnification, scanning speed, user friendliness and diagnostic functionality [148]. Validation studies on the application of IA in pathology show generally good agreement between glass slide and digital diagnosis [147,149,150]. Companies like Aperio and Leica have paired with microscope manufacturers to offer entire digital pathology packages that combine a microscope with computerized image acquisition and analysis [148,150,151]. There is great potential to improve patient diagnosis through efficient long-distance consultation between medical experts. However, better internet data security is required before this can occur.

Scan time is prohibitively slow on some IA systems. Also, the image from a scanned 20X or 40X slide is a large file requiring a large amount of storage space. Because of this, digital IA is more commonly used in medical education, where preselected “ideal” slides are used for lectures and image study banks. Currently, 33% of American medical schools teach using digital microscopy and this number is projected to increase to 50% within five years [149,150]. However, advances in technology are promising more rapid scanning of slides and more efficient image compression to reduce image size for storage and transmission.

IA is widely used in research to optimize an image and analyze stain intensity with

pixel counting software[138,152] Each image is scanned and appropriate areas screened for IA analysis. Each individual data file is exported and combined into a master file for sorting and statistical analysis.

1.4.2 Algorithms and automated IA with IHC

Advanced digital imaging systems pair the capture of whole slide digital images with diagnosis-aiding IA software. The Aperio Scanscope CS system (used in this project), incorporates computer algorithms that can recognize cellular structures and measure intensities of pre-calibrated colours in an image, effectively allowing the computer to "grade" the staining intensity of an image as a pathologist would [142]. This has advantages over manual pathology by removing observer bias and allowing calibrated, consistent diagnosis. When combined with reproducible stain quality, IHC IA can measure differences in staining intensity that are not distinguishable by the human eye [148,150,151].

For optimal IA, it is important to capture the image for analysis with a three charge-coupled device (CCD) camera which measures red, green and blue (RGB) separately. Older, single CCD cameras simulate all colors simultaneously, which may create measurable image artefacts. Because of the complexity of three CCD images, it is important that IA analysis of multiple samples is applied to images obtained from the same calibrated instrument. This removes subtle alterations due to light source, colour temperature, colour-correction and unique circuitry to optimize image quality. If this is not done, small differences could greatly affect IA measurement of faintly stained samples, creating erroneous results [32].

When properly applied, IA is unbiased and consistent. For experimental IA of stain intensity, samples are processed together with a known, quantified control sample. This allows correction for small changes in staining intensity between different experiments. The ideal reference is quantified and can be used to calibrate quantifying algorithms. Unfortunately, tissue heterogeneity prevents identical IHC controls. As a result, IA of IHC cannot be as stringently applied for quantitative measurement. To compensate for this in research applications, change in overall intensity is measured by averaging the measured intensity of a number of different control samples. By this method, only

overall increases or decreases in staining can be observed. This is why IA is only used in the semi-quantitative interpretation of clinical IHC stains[150]. True quantification of IHC results with IA has limited value with current methods and controls.[153].

1.5 Reverse transcription, quantitative real-time polymerase chain reaction (RT-qPCR) analysis

RT-qPCR was pioneered in the 1990s and has been used in research since 2000 [154–156]. RT-qPCR is currently used to quantify gene expression in samples of cells or tissues [157–160]. The amount of protein expressed in a sample is measured indirectly by converting the sample's mRNA into cDNA, which can then be amplified and quantified via PCR. An mRNA strand is reverse transcribed into its complimentary DNA strand using the enzyme reverse transcriptase. The DNA is then amplified using PCR. This allows a measure of gene expression, which is often assumed to reflect or correlate with the amount of protein expressed in the sample [155]. In this way, RT-qPCR provides additional, indirect measure/information about the transcription of genes within a sample. A positive correlation of RT-qPCR data with IHC-stain results shows that both mRNA and translated proteins/epitopes in the FFPE samples are relatively stable. It should be noted that the decay kinetics of mRNA and proteins due to the IHC procedure are not yet known for FFPE samples [155,157,158,160].

1.6 Research objectives

Premade, affordable, calibrated controls are currently not available for clinical use in IHC or for PT by EQA programs. This work explores different approaches to develop calibrated, standardized controls for diagnostic IHC QC and for EQA PT that could be affordable in daily practice. It also evaluated feasibility of IA for interpretation of control results in both types of applications.

1.6.1 Hypothesis 1: IA can be used in diagnostic IHC PT for EQA to rank individual laboratory performance, with respect to reference laboratory performance, using HScore and Laboratory Score: Reference Score Ratio (LSRSR).

Current cIQc EQA results are evaluated by a panel of pathologists. These evaluation results are used to calculate the sensitivity, specificity and the percentage of

correctly stained tissue cores for each laboratory. Cohen's Kappa values (κ) are calculated to evaluate for interlaboratory agreement of the participating laboratories *versus* the designated reference results. Acceptable Laboratories are ≥ 90% in agreement with cIQc reference results, with κ ≥ 0.8 (near perfect agreement) [161]. In this study, both IA and expert assessment were applied to evaluate the results of the PT by the cIQc. This study assessed whether relative scoring ratios like HScore or LSRSR can be used to standardize ER, PR and HER2 IHC instead of absolute pass/fail measurements. IA-based ratios may provide more useful information from a smaller number of samples than current TMA expert assessment methods.

1.6.2 Hypothesis 2: Cell line-based calibration models are applicable in IHC PT and can be used in place of current TMA models.

Calibration controls for an individual IHC laboratory method should be as similar as possible to the samples measured with these methods. Ideally, the processing of controls is identical to the processing of patient samples, using the same fixatives and fixation times. Formalin fixation (with or without decalcification), paraffin embedding and HIER are major components of tissue processing before it is analyzed by IHC [101]. To improve current EQA and PT methods, a cell line microarray (CLMA) was created containing nine human breast cancer cell lines. This CLMA was compared to currently used TMA calibration models made from human tumor samples in ER, PR and HER2 IHC PT. This work compares the performance of cell line samples to the performance of human tissue samples when used in IHC PT. IA was used to evaluate the performance of these samples.

1.6.3 Hypothesis 3: Cell line-based calibration models are applicable as daily positive controls for Class II IHC tests and can be monitored by IA with Levey Jennings-QC charts.

QC chart analysis has traditionally been applied to liquid-based laboratory testing. QC chart analysis has not, to date, been applied in diagnostic IHC laboratories. This work explores the feasibility of using IA to monitor and evaluate current clinical controls for ER, PR and HER 2 IHC made from human tumor and benign tissues. Current human, tumor-based controls from four major Canadian academic laboratories (that are reference laboratories for the cIQc Program) were analyzed with IA. New cell-line based

controls were also analyzed with IA. Levey Jennings-QC charts from tissue-based controls were evaluated. QC charts from tissue-based controls were compared to QC charts from cell-line based controls for ER, PR and HER2 IHC.

2.0 MATERIALS AND METHODS

A detailed list of chemicals and reagents is included in Appendix A. A detailed list of solution and medium recipes is included in Appendix B. A detailed list of instruments and relevant materials used is included in Appendix C. A detailed list of computer programs used is included in Appendix D.

2.1 Cell culture

Nine human breast cell lines were selected for inclusion in the breast marker CLMAs: AU-565, BT-474, HBL-100, HS-578T, MCF-7, MDA-MB-231, MDA-MB-435, SK-BR-3, and T47D. These cell lines have been demonstrated to express varying levels of ER, PR and HER2 [162–164]. All cell lines are available from the American Tissue Type Collection (ATTC) cell repository in Manassas, VA, in the United States. Cell lines were graciously gifted by Dr. Anderson, Dr. Carlsen, Dr. Lukong and Dr. Decoteau at the University of Saskatchewan, Saskatoon, SK.

All nine cell lines were adherent, growing in a monolayer attached to the bottom of a specially treated culture plate or flask. Cells were cultured in T25 (25 cm^2) and T75 (75 cm^2) vented culture flasks, with appropriate medium, in a 37 °C humidified growth incubator containing 5% CO_2. All media were supplemented with fetal bovine serum (FBS), penicillin and streptomycin. Some medium formulations had additional growth factors. All manipulations and medium preparations were performed in the sterile environment of a biosafety cabinet using aseptic technique and sterile pipettes. A brief summary of cell culture method is described below.

Cell lines were stored frozen in liquid nitrogen tanks in medium containing 10% (v/v) FBS and 5% (v/v) dimethylsulfoxide. Using aseptic technique, 1 mL volumes of frozen cells were thawed in a 37 °C water bath and gently transferred into a T25 flask containing 8 mL of pre-warmed medium. Following overnight incubation, culture medium was gently aspirated and replaced with fresh, pre-warmed medium. Healthy cell cultures were passaged a minimum of two times after thawing, before being used, so there was no residual affect remaining from being frozen. Cells were then subcultured into larger T75 flasks to grow up the large quantity of cells required. Healthy

cells were also refrozen for future use as required. Following initial storage, a vial of cells from each frozen culture batch were thawed and cultured to confirm viability.

Adherent cell lines grow to confluency cover the bottom of the flasks and cultures will cease growth and die once they reach confluency if they are not subcultured. To passage cells into new flasks, the culture medium was aspirated from the flasks with sterile pipettes. The cells were gently rinsed with 5 mL of pre-warmed, sterile, calcium- and magnesium-free Dulbecco's phosphate buffered saline (D-PBS). D-PBS was removed and 3 mL of sterile, pre-warmed 0.25 mM trypsin-EDTA solution was added. After a 1 min incubation at room temperature, all liquid was removed and the flasks of cells were placed in the CO_2 incubator at 37 °C for 2 to 5 min until cells rounded up and detached from the bottom of the flask. Then 6 mL of fresh, pre-warmed medium was pipetted into the flasks and the cells were slowly pipetted up and down. This was done to break up clumped cells and create a uniform seeding solution for sub-culturing new flasks. A portion of this cell solution was placed in a new T75 flask and brought to a final total volume of 20 mL with pre-warmed medium. Each cell line was diluted according to their optimal subculture dilution. Flasks were placed back into the 37°C CO_2 incubator. Medium was exchanged every 48 hr until cells again reached a sub-culturing density. This was repeated until sufficient cells were grown to make cell line blocks. To maintain culture robustness, cells were only passaged 20 times. Table B1, in Appendix B, shows the optimal medium and passage conditions for each cell line used.

Cells were cultured by a secondary method to confirm that IHC staining was observed in non-paraffin embedded cell samples. This was done in two ways. Cell lines were cultured directly onto glass slides for IHC staining using EZ-Slides. EZ-Slides have four or eight sterile, separate culture chambers per slide. Cell lines were grown on EZ-Slides with the same methods, medium and growth conditions as in flasks. When cell lines reached confluency, culture medium was poured off and the cell chambers removed form the slide according to manufacturer instructions. Cells that remained attached to the slides were fixed for IHC staining, as described in Section 2.3.

Three cell lines (AU-565, MDA-MB-231, SKBR3) adhered poorly to EZ-Slides and washed off during the IHC-staining process. These cells were cultured with standard methods in T25 flasks and collected in the same manner as for cell pellets (described in

Section 2.2.1). The D-PBS-rinsed cells were spun down onto positively charged glass slides using specialized funnel filters and a Cytospin 4 centrifuge. A total of 2 X 10^6 cells were spun onto slides at 2000 RPM for 3 min. Cells were then fixed for IHC staining as described in Section 2.3.

2.2 FFPE cell blocks

2.2.1 Creation of cell pellets

It is ideal to have control material that is directly comparable to patient tissue samples regarding tissue (or cell line sample) processing, including the type of fixative, time of fixation, and embedding medium and methods of embedding. Cell lines were collected and made into thrombin pellets, as described briefly below. This step was done in order to aggregate the cell lines in clotted plasma and create a sample that could be processed with the same methods as patient tissue samples. Following this, the cell pellets were fixed in 10% buffered formalin and embedded in paraffin in the same manner as patient tissue samples at Saskatoon City Hospital (SCH) of the Saskatoon Health Region (SHR), also as described below.

Trypsin, which is used to subculture adherent cells, can cause minor changes to cellular morphology during the detachment of cells from the culture vessel [165]. Because of this, trypsin was not used when harvesting cells. EDTA alone causes adherent cells to detach as a layer rather than rounding up individually. Removal of cells at passage density (as per Table B1) was carried using sterile D-PBS containing 2 mM EDTA. A brief description of this method is described below. Growth medium was removed from the flasks. Cells were rinsed with 10 mL pre-warmed, sterile D-PBS. This was removed and 10 mL of room temperature, sterile D-PBS + EDTA was added to the culture flask. Flasks were incubated in the CO_2 incubator at 37°C for 3 - 5 min until cells detached. Flasks were tapped firmly twice on the laboratory bench to completely detach the cells into solution. Cells were gently pipetted up and down twice with a 10 mL sterile pipette to break up the cellular monolayer and the re-suspension was then placed into sterile 15 mL centrifuge tubes. Cells were pelleted by centrifugation at low speed (1000 RPM) for 2 min. The EDTA solution was removed and cells were rinsed by resuspension in sterile D-PBS and centrifuged again. Before centrifugation, a sample was taken to quantify the

cells with a haemocytometer according to standard methods[165] (Section 2.2.2). The D-PBS was poured off and a final D-PBS wash was carried out in the same manner. Cells were combined or divided at this stage to obtain the required number of cells per pellet (See section 2.2.2) in each tube. The final rinse solution was poured off and a P-100 Gilson pipettor was used to carefully remove any remaining D-PBS.

The cells were resuspended in 150 µL of human AB plasma with gentle finger vortexing. An equal amount of thrombin solution from the Beckman Coulter HemosIL Thrombin Time kit was added. The tube was finger vortexed for 10 sec and placed in a rack for 3-5 min at room temperature to allow a firm clot to form. Then 5 mL of 10% buffered formalin solution was added. Inversion was used to loosen the clot from the tube for complete fixation.

2.2.2 Cell pellet optimization

Due to the costly nature of cell culture, it was important to determine the appropriate number of cells required for a diagnostically useful cell pellet. A number of blocks were created with different cell concentrations. Once cells were harvested, rinsed and resuspended in D-PBS, they were counted in a Springline haemocytometer, by using a mixture of 100 µL of cell solution with 100 µL of 0.5% trypan blue. [165].

Once counted, the volume required for a specific number of cells in each cell block was suspended in a thrombin pellet. Thrombin pellets were fixed, embedded, sectioned and IHC-stained. Stained slides were examined manually to ensure proper dispersion of the cells in the pellet. Slides were also scanned and analyzed by digital IA to ensure the IA software could correctly measure staining in cell pellet blocks. The final thrombin cell pellets used for IHC and IA contained 2.0 X 10^7 cells in a 300 µL volume. This number of cells was present in two 95% confluent T75 flasks. For cells harvested at 70-75% confluency, three T75 flasks were required, while cells harvested at 50% confluency required six T75 flasks to be used.

2.3 Fixation and embedding of tissue and cells for histological staining

Fixation is used to prevent post-mortem decomposition, thus preserving structure. Fixation prevents autolysis and prevents cells from changing their volume or shape. As previously discussed in Section 1.2.1.1, there are optimal times for fixation of

specimens in buffered formalin before the samples are further processed and embedded.

Cells pellets were fixed overnight (12-14 hr) in 10% buffered formalin. Clinical tissue samples are fixed for 8-48 hr in 10% buffered formalin. After fixation, pellets were placed into pre-labelled plastic cassettes. Smaller biopsies, tissue samples, and cell pellets are placed inside a folded piece of histological-grade tissue paper before being placed inside the cassette to prevent sample loss during processing. The thrombin cell pellets form in a conical bottom tube, so the pointed tips of the cell pellet were trimmed off the narrow end of the pellet with a sterile razor blade. This tip and the remaining disc-shaped clot were placed horizontally within the paper and cassette to create the largest diameter sample possible. In this way, the sample spans the entire depth of the finished paraffin block. This allows the maximum number of samples cores per fixed cell pellet and sections per finished block. Up to three cell pellets were placed in each cassette to optimize cost and labour. Cassettes were placed back into 10% buffered formalin until processing later the same day.

Samples were processed for embedding at SCH using a Sakura Tissue Tek TEC 5 Tissue Embedding Console. The SCH standard fixation program was used, which takes 12 -16 hr to complete. To dehydrate the tissue, the processor immerses the cassette into multiple baths of progressively concentrated ethanol with final rinses of 100% ethanol. The clearing stage then replaces the ethanol with a number of 100% xylene rinses. Upon completion, the dehydrated tissues are miscible with paraffin. Processed tissue is taken out of the cassette by a technician and set in a metal mould containing a small amount of liquid paraffin. After the sample is levelled in the mould, additional paraffin is added on top to create the final paraffin block [166,167].

After some experimentation, optimal fixation for cytospin and EZ-slides was determined to be 30 min in 10% buffered formalin. After formalin fixation, cell culture slides were placed into sterile PBS and immediately brought to the SCH IHC Laboratory for staining.

2.4 TMA assembly

TMAs provide a way to compile multiple tissue samples that were processed independently of each other (at one or more laboratories) into one paraffin block. This composite block can be sectioned onto slides to simultaneously evaluate these samples together [168]. Microarrays are constructed by acquiring a small punch core sample from areas of interest in FFPE samples. Cores are incorporated into a recipient paraffin block and heated to permanently bond the block together. Construction of duplicate blocks, or blocks with multiple sections of the same tumor in it, allows for immediate comparison, and large savings, in research that was not previously possible [168–170]. Core samples are commonly sampled as 0.6 mm, 1 mm or 2 mm cores, depending on the area of interest, the type of lesions being evaluated, and the type of analysis performed on the block. TMAs are optimally suited for IHC IA analysis because the selection of the areas for analysis is included as part of the sample core collection [171–173].

2.4.1 Sample selection and core collection

The first step in TMA design is the selection of samples for inclusion. In this study, microarrays created from FFPE cell lines were assembled by myself and, in the case of tissue controls, by the histology laboratories of SCH (Saskatoon, SK), Vancouver General Hospital (VGH) (Vancouver, BC) and Jewish General Hospital (JGH) (Montreal, QC). Searches of databases with records relevant to quality assurance operations were performed to ensure that FFPE breast tumor blocks with varying levels of expression of ER, PR and HER2 were selected. Data search for QA purposes is exempt from Research Ethical Board review.

Areas of interest in the FFPE blocks were identified from ER, PR and HER2 IHC-stained slide sections of each block. Punch cores (1 mm and 2 mm) were taken from the selected areas using the donor punch of the Beecher TMA instrument. Two to three cores were taken from each block, depending on the tumor size. Cores were stored in labeled 0.5 mL microcentrifuge tubes in a cool dark place until assembled into a finished array at a later time. A sample bank of cores was created in this manner with the identity of samples kept confidential. Records linking patient data to de-identified samples were kept in a secure location separate from TMA information.

2.4.2 Block assembly

A recipient block was created for each array by standard methods [168,173]. Paraffin cores were removed from a blank paraffin block using the recipient punches of the Beecher TMA instrument. A linear distribution of 1 mm was left between each hole. The 1 mm punch was damaged partway through this study and the 2 mm punch was adopted to complete the study. Cores were placed into the recipient block using the flat end of a stainless steel spatula, which was carefully cleaned between samples. The identification and location of each core was recorded. When completed, the block was placed paraffin-side down onto a clean glass microscope slide and heated to 40°C for 20 min. Following this, the block was firmly and evenly pressed down onto the glass slide. The block and slide were immediately placed onto a block of ice for 30 min to cool the paraffin block evenly. Once cooled, the slide was gently removed from the completed array [173].

2.4.2.1 Microarray blocks used in this project

A number of array blocks were assembled for this project. They are described in detail below.

cIQc Run 13 TMA and CLMA: cIQc Run 13 TMA was a neoplastic and benign human TMA assembled at a cIQc Program Laboratory (University of British Columbia, Vancouver, BC). This block, referred to in this study as “Run 13 TMA” contained 56 tissue cores: six rows of nine cores each, with two control alignment cores of FFPE human liver tissue added to the top left of the block. Nine FFPE cell line samples for Run 13 were placed into a separate CLMA block. This block was used in a cIQc EQA run for ER, PR and HER2.

Special Run TMA/CLMA: An array block, hereafter referred to as “Special TMA” was assembled in our laboratory. This array contained 47 cores. There were five rows of nine cores, with two alignment cores of FFPE benign control liver tissue on the top left of the block. The first two rows were 2 mm cores of all nine cell lines. The remaining 27 tissue cores were 1 mm tissue cores of FFPE breast cancer blocks selected and

donated to cIQc by JGH. This block was used in a cIQc EQA IHC run for ER, PR and HER2.

Final CLMA: For the CLMA used for PT in this project, a search of an internal SHR database of ER-stained, PR-stained and HER2-stained breast cases from 2005-2009 was performed. A total of 110 cases were selected with varying levels of expression of ER, PR and HER2, with tumor sections large enough to get two or three 2 mm cores from each block. Needle biopsies, cases originating from external hospitals, rare cases and tumors smaller than 1 mm were excluded. From this initial group, 22 FFPE case blocks were selected. This array, referred to hereafter as “Final CLMA”, contained 38 2 mm cores. The Final CLMA was a six by six grid of cores, with two alignment cores on the top left of the block. Two cores of each of the nine cell lines (18 total) and 22 cores of tissue were included in this block. This was the final array created for this study.

JGH CLMA: JGH assembled a HER2 IHC control array containing 16 FFPE breast tumor samples. This array also included two FFPE cell lines created for this project (AU-565 and MDA-MB-231). Four tissues cores were selected from each HER2 Score (0 to 3+). Cell line AU-565 was included because it has strong HER2 IHC staining (HER2 Score = 3+). Cell line MDA-MB-231 was included because it is negative for HER2 IHC staining. Sequential slides from this block were used as HER2 IHC controls at JGH. The HER2 control slides were sent to me for IA once JGH analysis with them was completed.

2.5 IHC

This section gives a description of the IHC-staining process, from the sectioning of the block onto the slides through the pre-analytical, analytical and post analytical stages to the finished slide.

2.5.1 Sectioning

FFPE block sections were cut at SCH on a Leica RM2235 microtome. Four-micron thick sections were floated onto positively charged microscope slides using a distilled water-filled, warmed tissue bath with no additives. Excess water was then blotted from

the slide, allowing it to air dry. For optimal staining, slides that would be stained within 2-4 days of cutting were stored at 4°C, while slides stored for a longer time were kept at -80°C [174]. Prior to IHC, chilled slides were thawed and completely air-dried (if required) for 2-3 hr before de-paraffinization.

2.5.2 De-paraffinizing and rehydration

It is important to remove all the paraffin from the sample for reproducible and complete IHC staining. Initial baking of the slides at 60 °C for a minimum of 1 hr (maximum overnight) begins this process. Once the deparaffinizing and rehydration process begins, the slides cannot be allowed to dry out at any time or the resulting IHC stains may have high background and non-specific, uneven staining [36].

To remove paraffin, the baked, cooled slides were placed into a TissueTek II slide tray. The slide tray was placed in a series of solutions, with a vigorous agitation at the start of each step. Four 5 min xylene washes began the process to remove remaining paraffin from the sections. This was immediately followed by 3 X 10 sec 100% ethanol washes, 2 X 10 sec 95% ethanol washes and a 1 min 70% ethanol wash. Following this, the slides were placed thoroughly rinsed with numerous distilled water rinses to remove all ethanol. Rehydrated slides were placed in distilled water or PBS, and processed with HIER [36,87].

2.5.3 HIER

For FFPE tissues, it is usually necessary to pre-treat the tissues with HIER in order to de-mask epitopes for optimal staining (Section 1.2). HIER is optimized for each tissue type and antibody [175]. For ER, PR and HER2 IHC, SCH uses the following optimized HIER method. De-paraffinized, rehydrated slides were placed into a TissueTek II slide tray containing freshly prepared 1 mM EDTA (pH 9.0). Any empty places on rack were filled with blank glass slides to ensure consistent heating. To prevent explosion during heating, the lid was loosely placed on the full buffer chamber containing the slides.

The slides were heated in an 1100 W microwave for 13 min at 60% power. After this, the slides were left in the hot buffer for another 20 min before rinsing briefly with room temperature distilled water. Once the slides cooled to room temperature, they

were thoroughly rinsed with running distilled water for 2 min to remove all EDTA [36,47,52]. Slides were kept in distilled water until they were loaded onto the IHC autostainer.

For cytospin and cell culture slides, the fixation process is less harsh, requiring a less harsh HIER procedure. Standard HIER at SCH for cytospin slides is to place them into pre-boiled EDTA HIER buffer for 15 min before cooling and rinsing with distilled water prior to IHC. Alternatively, FFPE cell line slides (cut from blocks containing cell pellet material processed the same as human tissue) received the same HIER as regular biopsy samples.

2.5.4 Immunostaining

All IHC slides for this project were stained at SCH on a Dako Autostain Plus IHC instrument with Dako Envision Plus reagent kits. Specific Envision Plus kits were paired with their appropriate primary mouse, goat or rabbit antibodies. The autostainer has a hydration chamber to maintain humidity and prevent the drying out of slides during staining. Manual staining is done with the aid of a small hydration chamber for incubations that prevents dehydration. A summary of the IHC methodology and procedure follows.

De-paraffinized HIER-prepared slides were marked with a Dako hydrophobic marker on either end of the slide to help contain the stain solutions on the slide surface. Slides were then loaded onto the autostainer and Dako IHC buffer was placed on top of the slide until the procedure began. A standard IHC protocol was entered into the Dako Autostainer software program. This protocol begins with a 5 min block step to remove endogenous peroxides, followed by a buffer rinse. The next step is a 30 min primary antibody incubation followed by two buffer rinses. The next step is a 30 min secondary antibody incubation step followed by two more buffer rinses. Final 10 min DAB chromagen incubation is followed by a distilled water rinse. The Envision Plus kits contain the peroxidase block solution, polymer HRP-labeled secondary antibody and a DAB chromagen solution. A total of 300 µL of all solutions were used for each slide stained.

It is SCH policy to have sections of control blocks mounted onto all sample slides before staining. Controls are specific for the antibody used. The monoclonal primary

antibodies used in SCH IHC-staining of patient samples have been validated for use. Optimal dilutions are re-evaluated with every batch received. During this study, rabbit anti-human ER (α-subunit) antibody (Thermo Scientific, Clone SP1) was diluted to 1/50 before use. Mouse anti-human PR (A-isoform) antibody (Novocastra, Clone 16) was diluted to 1/100 before use. Rabbit anti-human HER2 antibody (Thermo Scientific, Clone SP3) was diluted to 1/250 before use. Antibodies were diluted with Dako antibody dilution buffer.

2.5.5 Dehydration, counterstaining and coverslipping

After IHC, the slides were placed in distilled water to be counterstained and coverslipped. DAB staining was enhanced with a 5 min incubation in 2% (w/v) copper sulphate solution (made fresh weekly). Slides were rinsed under running room-temperature tap water for 5 min. Following this, unstained tissue was counter-stained with 15 dips into light Hematoxyllin 1. Slides were thoroughly rinsed with room-temperature tap water. To optimize the blue counterstain, the slides were then placed into freshly prepared alkaline water (200 mL distilled water + 0.5 mL ammonium hydroxide) for 30 sec and then rinsed well with running room-temperature tap water.

Slides were dehydrated in a fume hood using a TissueTek II tissue stain rack. Slides were placed in a series of ethanol washes (~10 sec per rinse, 10 times vigorous dips up and down). Three washes of 95% ethanol were followed by three washes of 100% ethanol to remove all water. A final four washes of xylene removed all traces of ethanol. Slides were stored briefly in xylene until they could be covered with 20 mm X 40 mm coverslips and Permount mounting medium.

2.5.6 Interpretation of results

Current ASCO/CAP guidelines used for the scoring of ER, PR and HER2 IHC results were discussed in the Introduction Section and in more detail immediately below [7,20]. The guidelines basically apply to visual inspection by an experienced pathologist using a light microscope, but also include guidelines for scoring by IA. HScore value can be derived from visual inspection and counting of cells, but is much faster to do derive

HScore using IA. An example of nuclear (ER, PR) and membranous (HER2) IHC staining of FFPE tissue and cell lines can be seen in Figure 1.

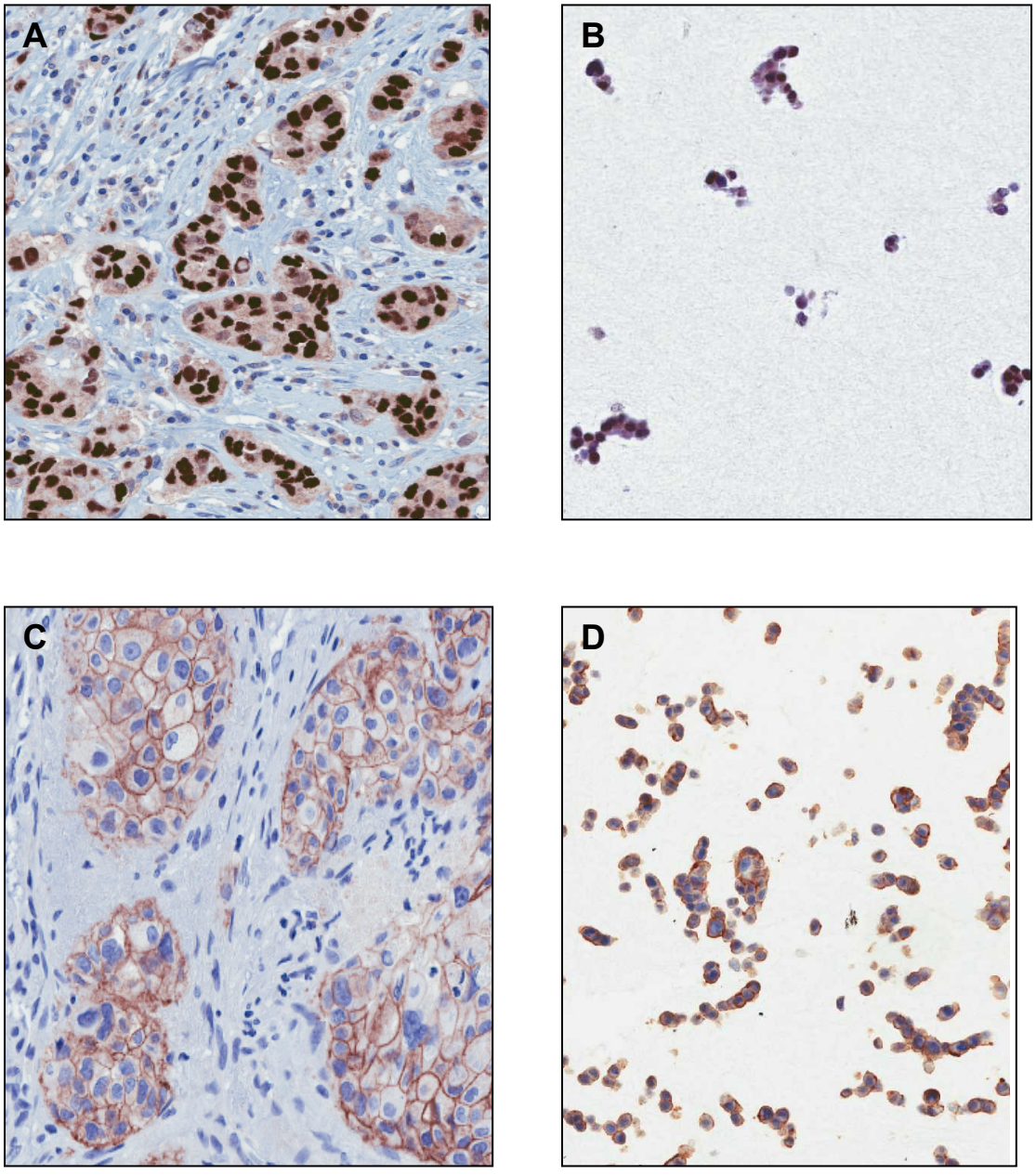
It is important to note that the expert assessment by cIQc Program differs from the above guidelines and is limited to establishing whether the PT samples are positive or negative for a specific marker. These expert-assessor scoring results are used to calculate a participating laboratory's level of IHC-stain sensitivity, specificity and κ in relation to a designated reference value. In this project, the pass/fail rate of participating laboratory the performance was compared to their success rate, which was observed using HScores calculated from IA algorithm results. HScores were used for LSRSR calculations (Section 2.6.3.4) and the creation of QC control charts (Section 2.6.3.2). Because the methods of scoring are different, the cut off points regarding the pass/fail rates are also different for each method. The cIQc Program expert assessment pass/fail cut off was at 90% sensitivity, with specificity and κ cut off set at ≥ 0.80 (near perfect agreement). The IA/LSRSR pass/fail cut off was when participating laboratory LSRSRs fell outside of the reference LSRSR range for multiple sample measurements.

2.5.6.1 ER, PR and HER2 IHC scoring

Estrogen and progesterone receptors are located in cell nuclei. Irrespective of the method used, IHC staining is deemed to be “positive” for ER and PR if greater than 1% of the total nuclei are at least weakly positive. Both intensity of staining and the percentage of positive cells are assessed. For ER and PR, positivity is defined as a change in color of the nucleus of the tumor cells from blue (negative) to gray (weakly positive) or brown (moderately and strongly positive).

HER2 is a membranous receptor. As discussed in section 1.1.3, a HER2 score of 3 indicates a clinically positive sample. HER2 score of 2 is defined as equivocal and it is obligatory to study these tumors by FISH, to confirm if the sample is considered a true positive or not [9,10]. A HER2 score of 1+ and 0 are applied to clinical practice as negative, and the patients with such tumors do not qualify to receive targeted therapy. Although such scoring is used in clinical practice, the focus of this work is not to stratify patients for therapy, but rather to address technical aspects of variation in calibration of the methods. Therefore, the above semi-quantitative scoring was not used because it is

Figure 1: Nuclear and Membranous IHC-Staining of FFPE Tissue and Cell Line Cores. Brown color is positive staining for protein of interest. Blue staining is counterstaining of negative tissues with hematoxyllin. A, tissue, nuclear IHC stain; B, cells, nuclear IHC stain; C, tissue, membranous IHC stain; D, cells, membranous IHC stain.



not suitable for calibration due to a very narrow value range. For this reason, HScores of HER2 IHC results were used in this research. Monitoring of calibration is essential to decrease variation in the system and provide reproducible results, which can then be interpreted on any clinically acceptable scale for patient treatment.

2.5.6.2 Linearity of IHC staining

IHC staining is a cumulative process, with primary antibodies binding to epitopes of interest, followed by secondary, HRP-labeled antibody binding to the primary antibody. The Dako Envision Plus reagent kit used in this project uses a secondary antibody that is bound to a polymer, which is in turn bound to a large number of HRP molecules. This polymer-HRP-bound secondary antibody is designed for intense staining efficiency. The Lambert–Beer law usually used in determining the linear useful linear range of a chromagen can only be applied to pure absorbing compounds. The DAB chromagen used in IHC, exhibits a scattering behavior, and is spectrally identical to natural pigments like melanin [32]. Luckily, a reasonable dynamic range of measurement can achieved using DAB together with IA algorithms that are optimized for DAB. Other red-based stains that show better linear ranges for IA measurement are not commonly used in clinical IHC [176].

Currently used IHC detection systems are designed to amplify the observed staining signals. This amplification results in a logarithmic type of association with the concentration of target epitopes. However, lower levels of protein/epitope and higher antibody dilutions have been shown to result in a more linear range of staining, which allows for quantitation of the epitope in the tissue. In contrast, more highly expressing tissues will show less variation in the intensity of the signal because the system oversaturates at higher expression levels. Therefore, positive controls with weak expression of epitopes are recommended for both manual and IA systems [10,21,22,177].

2.6 IA

IA can reduce human error in IHC interpretation [32,139]. This project examined the application of IA in IHC for use in EQA and PT. IA provides a non-subjective reference/scoring method for evaluating the staining performance of one or more laboratories [161]. IA was also applied to daily QC for the interpretation of in-laboratory,

on-slide controls to evaluate run-to-run performance. A combination of digital IA evaluation of IHC of ER, PR and HER2 with cells lines as positive controls and HScore/LSRSR evaluations were applied.

2.6.1 Aperio scanner

The Aperio Scanscope CS system used in this project is a powerful three CCD quantitative camera that captures 20X digital images from five slides at a time, using a high quality microscope. Once the digital image is saved to a server, the Aperio Spectrum (version 11) software works in consort with Aperio Imagescope viewing software (version 10). Through Spectrum, appropriate algorithms were applied to digital images to identify cellular structures and measure DAB stain intensity. Aperio TMALab software (version 10) was used for scoring and analysis of TMA slides using the same algorithms. Numerical results were exported in Excel 2007 format for further analysis.

2.6.2 Algorithms for assessment

The algorithms used in this project have been optimized by Dr. Emina Torlakovic for use in the cIQc EQA Program. This optimization ensured that the algorithm results correlated with clinically used cut off points. IA algorithms are calibrated using measurements of the RGB values of the positive DAB (brown) and negative hematoxyllin (blue) stains used in the SCH IHC Laboratory. This calibration reduces small variations in the measured colour spectra that can have quite dramatic effects on the algorithm-calculated values in higher staining samples [32].

These calibrated RGB values were used for all slides in this project. Stain intensity value ranges for 0 (negative), 1+, 2+ and 3+ (highest intensity) were defined by Aperio in the their algorithms. These algorithms can identify individual cells and cell nuclei, but cannot differentiate tumor from normal tissue. Specific areas of interest on a slide must be pre-selected for IA to ensure appropriate analysis. Algorithms can be applied to a whole image, a selected annotation of an image or individual cores of a TMA.

Two Aperio algorithms (specifically modified and calibrated as described above) were used for this project. “ET Nuclear IHC new colours” measures nuclear DAB staining intensity for ER and PR IHC. “ET IHC Membrane for RUN5” measures completeness of membrane staining and DAB membranous stain intensity for HER2

IHC. These algorithms use mathematical formulae to identify nuclei and membranous structures. The RGB stain intensity of each region is measured, and the cells are classified into the appropriate stain intensity. The total number of measured cells, nuclei and complete membranes are determined. The percentage of total cells of each intensity (0 to 3+) is then calculated. A virtual annotation image overlay can be created during analysis to allow the user to visually determine that IA has been applied appropriately. This overlay uses a different colour for each intensity range: Red for 3+, orange for 2+, yellow for 1+ and blue for 0. Detailed algorithm settings are listed in Supplemental Table 1.

TMALab software allows for a TMA map to be created for IA. Each slide can be manually mapped to a template. Once all slides and cores are manually identified, each core can be analyzed individually with any algorithm. TMALab was used in this project for IA of individual cores in all TMAs and CLMAs.

In IA, damaged or folded TMA cores can be erroneously scored. Overlapping and damaged tissues will be interpreted by IA colour measuring algorithms to be more intensely stained than they actually are. IA algorithms may also incorrectly record a missing TMA core as a negatively staining core. Visual inspection of slides and IA results are required to identify and exclude any improperly analyzed cores.

2.6.3 Statistical analysis

IA HScore results were statistically compared as indicated in the results (*t*-tests, F-tests, κ and Chi-squared analysis of variance). IHC-stain results and laboratory results in the cIQc EQA Program were graphically compared through QC charts, histogram plots and boxplots of HScores and LSRSRs. SPSS 19 and Excel (2007 and 2011) were used for all statistical analyses.

2.6.3.1 HScore

Completed algorithm scoring data is exported as an Excel compatible spreadsheet. IA data from each slide was exported as a separate file. Individual slide and core data was later manually copied and pasted together into summary worksheets for analysis.

HScores were calculated using the following formula [81]:

HScore = 3 x (percent 3+ cells) + 2 x (percent 2+ cells) + 1 x (percent 1+ cells)

IA HScores were used to derive LSRSR values as described in Section 2.6.3.4

2.6.3.2 Levey Jennings control charts

Levey Jennings control charts were created with a template created in Excel 2007. HScores calculated from the IA results of individual controls, analyzed over time, were pasted into this Excel template. A Levey Jennings control chart[112] places each daily measured control progressively on the X-axis of the graph. The HScore of each control sample is plotted on the Y-axis. Two and three SD from the mean of the HScore data set are calculated and plotted on these graphs, allowing for the visualization of any samples that fell outside of the acceptable performance range [112,113,115,133]. Some study results are expressed as Levey Jennings QC charts.

2.6.3.3 clQc PT performance - the use of Garratograms

In the clQc EQA Program, a group of expert pathologists and technologists evaluate ER, PR and HER2 IHC-stained slides submitted by participating laboratories. Scoring results are expressed as garratograms (to credit the contribution of John Garratt from Vancouver, BC who first used heat maps in EQA to illustrate concordance of results between participating laboratories). Garratograms allow for visual comparison of stain results of a reference laboratory or method to that of a participating laboratory. Different colors are used to designate results. In a garratogram, samples showing positive staining (>1% nuclear staining) are denoted with red squares, negatively staining samples are denoted with white squares and unacceptable or missing samples are denoted as yellow cores. When the results of all cores for all participating laboratories are lined up beside those for reference laboratories, any cores improperly stained are easily identified. Laboratories are considered to have acceptable staining method if they have ≥90% agreement or the same level of sensitivity and specificity with reference laboratory results. κ (a statistical measure of inter-rater agreement) is also calculated to

compare participant IHC stain scoring to cIQc stain scoring. A $\kappa \geq 0.8$ (nearly perfect agreement) is considered acceptable performance.

2.6.3.4 LSRSR calculation

This project applied LSRSR calculations as an indirect scoring method and a calibration tool to compare and measure performance of laboratories in the cIQc EQA Program through IA. LSRSR is calculated from HScore values calculated from IA data. To calculate LSRSR, the HScore of each participating laboratory is divided by the HScore of the reference or reference laboratory, which is used as a "gold standard". What is used as a reference result may vary; either a single reference laboratory's results or, alternatively, an average HScore of multiple reference laboratories may be used. As no definite true "gold standard" exists for calibration of IHC methods and results, a number of reference core samples and laboratories are used to increase the accuracy of the selected reference results in EQA.

In this study, the HScore of a participating EQA laboratory is divided by the average reference HScore of six selected reference laboratories to create a ratio. This ratio is the LSRSR for that laboratory for any specific tissue or cell line sample (tissue core or cell line core or any other sample selected for this purpose). LSRSR reflects the degree of agreement between the participant and the average reference result and, ideally, it is 1.00. LSRSR is usually calculated for the reference laboratories for a number of representative cores (four to six) to give the acceptable LSRSR range for each core (although any number of samples may be included by the EQA program for this purpose). The minimum and maximum LSRSR values from the six selected reference laboratories was the acceptable LSRSR reference range for participating laboratories for any sample in an EQA PT run. Although many tissue samples are included on the test slides, only a few selected tissue (or cell line) samples are used for LSRSR calculation. Sample selection is based on the reference laboratory HScore for that sample, which should optimally be in the informative, more sensitive, HScore range (50-100 whenever possible). HScores of participating laboratories are divided by the average reference HScore value for each of the selected informative tissue or cell line samples. LSRSR scores that fall within the reference LSRSR ranges indicate that

participating laboratory results are in the range of reference laboratories results and that the participant is performing in an acceptable manner for the particular EQA assessment.

The creation of an acceptable LSRSR reference range is best depicted graphically. The average HScore was calculated from the IHC HScores of the six preselected reference laboratories for each of the selected informative samples (either tissue cores or cell line cores). This average HScore was then used as the denominator to calculate an LSRSR for each reference laboratory for each informative sample. Reference laboratory LSRSRs are then plotted on the y-axis, with each informative sample as a point on the x-axis. When all LSRSR are plotted, the minimum and maximum reference LSRSR for all the informative samples are easily identified and any extreme outliers can be identified and excluded. The minimum and maximum LSRSR values from this graph become the acceptable reference LSRSR range for that specific IHC stain PT run. Extreme LSRSR values were identified and selected as outliers using a form of delete-1 jackknife resampling [178,179]. By this method, any graphed values that appeared to be outliers were identified. These suspect values were removed, one at a time, from the graphed data group. The acceptable LSRSR range was re-evaluated once this value was removed. If the suspected LSRSR was outside the revised acceptable LSFSR range, then the suspected value was discarded. This acceptable reference LSRSR range was then used in PT evaluation of all participating laboratories [161].

2.6.3.5 Boxplots and histograms

The closer a laboratory LSRSR value is to 1, the closer the IHC staining of that laboratory is to that of a reference laboratory or laboratories. Graphically plotting the LSRSR or HScore of a laboratory as a histogram and boxplot will visually represent a laboratory's EQA IHC performance in comparison to the entire test group. Poor performance is easily visually identified using these plots, making them a useful method of presenting EQA results to participating laboratories.

2.7 RT-qPCR

To more effectively use cell lines as IHC controls, it is important to observe the consistency of expression of ER, PR and HER2 in FFPE cell lines after processing

through FFPE and HIER for IHC staining. Formalin fixation and paraffin embedding may damage proteins of interest [160]. Decay kinetics of transcripts and proteins from the IHC procedure are not yet known for FFPE samples.

The laboratory at the University of British Columbia (Vancouver, BC), which prepares the majority of cIQc Program FFPE samples for EQA, generously arranged for RT-qPCR analysis of ER, PR and HER2 gene expression in the nine cell line blocks used in this project. Such transcriptional analysis of ER, PR and HER2 allows comparison with the protein levels observed via IA IHC. These RT-qPCR results should provide insight into the possible degradation of ER, PR and HER2 in the processing of my FFPE cell line samples during IHC staining. This knowledge should be helpful in the practical application of these cell line samples as IHC control samples [155,157,158,160].

The RT-qPCR analyses were performed at the Bernard Laboratory at the University of Utah using a previously optimized method [157]. RT-qPCR analysis to measure ER, PR and HER2 gene expression was carried out with mRNA extracted from the nine FFPE cell lines in this project. This was done using a Qiagen Total RNA extraction kit. cDNA was prepared with SuperScript III (Invitrogen) and a Qiagen DNA clean up kit. Two replicates of each sample were on a Roche LightCycler480 RT-qPCR instrument, using Roche Master Mix double-stranded DNA-specific dye [157,158].

Results were analyzed with Profiler software. Results are expressed by a relative quantification method [157], using an external efficiency curve in which the calibrator gene is valued at 10 ng. To correct for differences in sample quality and cDNA input, the normalized sample ER, PR and HER2 copy numbers were then adjusted relative to the results for five housekeeping genes. Adjusted sample results were then averaged to obtain the final relative RT-qPCR gene expressions [157,158].

A slide of a CLMA containing the cell lines was also cut and IHC stained for ER, PR and HER2 by cIQc reference laboratories. The IA-calculated HScores allow for comparison of IHC-measured ER, PR and HER2 expression with measured ER, PR and HER2 mRNA levels.

2.8 Collaboration

2.8.1 cIQc involvement

The EQA/PT component of this project was possible due to collaboration with the cIQc Program. As previously described, the cIQc Program created and sectioned the slides, and was responsible for the delivery of slides to participating laboratories. cIQc pathologists and technologists evaluated Run 13 PT slides and shared their garratogram scoring results. All cIQc Program participating laboratories were included in the final data evaluations. The cIQc Program selected statistically relevant numbers of reference material for EQA TMA assembly. The reference laboratories used for LSRSR calculations were selected by the cIQc program based on consistent previous EQA performance, from the initiation of the EQA program up until the present.

2.8.2 Laboratories involved

This project was completed in collaboration with a number of Canadian clinical IHC laboratories. The SCH performed all the paraffin embedding of my cell pellet blocks as well as the sectioning and IHC staining of these blocks onto slides. SCH was one of the cIQc founding IHC laboratories and continues to participate in the cIQc Program, where it has been graded as one of the reference laboratories. SCH performed the IHC-staining of ER, PR and HER2 slides from the cIQc CLMAs and TMAs of the cIQc Program. SCH also included Final CLMA slides into their daily IHC runs for my interpretation with IA and Levey Jennings QC charts. The FFPE blocks used to create my Final TMA were selected from a SHR breast cancer patient database. The tissue-based control slides for ER, PR and HER2 were selected from the SHR 2009 slide archives for analysis by IA with QC charts.

Jewish General Hospital (JGH) from Montreal, Quebec, another cIQc reference laboratory, contributed a number of FFPE breast cancer tissue samples used in the Special CLMA in this project. JGH also generously agreed to include two of my cell line blocks into one of their FFPE on-slide HER2 IHC control blocks. Once patient diagnosis was complete, JGH submitted this set of control slides to me for IA and QC chart analysis. JGH also generously donated sets of tissue-control slides for IA QC chart analysis of current in-laboratory tissue controls.

Vancouver General Hospital (VGH), Vancouver, British Columbia, is currently a core laboratory for the cIQc Program. Clinical and academic research laboratories, under the advisory of Dr. Blake Gilks and John Garratt, create and section the TMAs used in the cIQc EQA Program. The University of British Columba research laboratory arranged for RT-qPCR analysis of the nine cell line FFPE blocks created for this project at the Bernard laboratory at the Huntsman Cancer Institute of the University of Utah, in the United States.

Two laboratories that regularly participate in the cIQc Program, from Edmonton, Alberta and Toronto, Ontario have made available several sets of ER-, PR- and HER2-stained on-slide IHC human tissue controls to this project. These sets of tissue-control slides were used for IA and QC chart analysis in this project.

3.0 STUDY RESULTS

The three hypotheses of this study were addressed in a number of PT and QC experiments. The detailed data relevant to each hypotheses are presented in this section.

3.1 Introductory work

3.1.1 Cell line block creation, slide scanning and IA evaluation

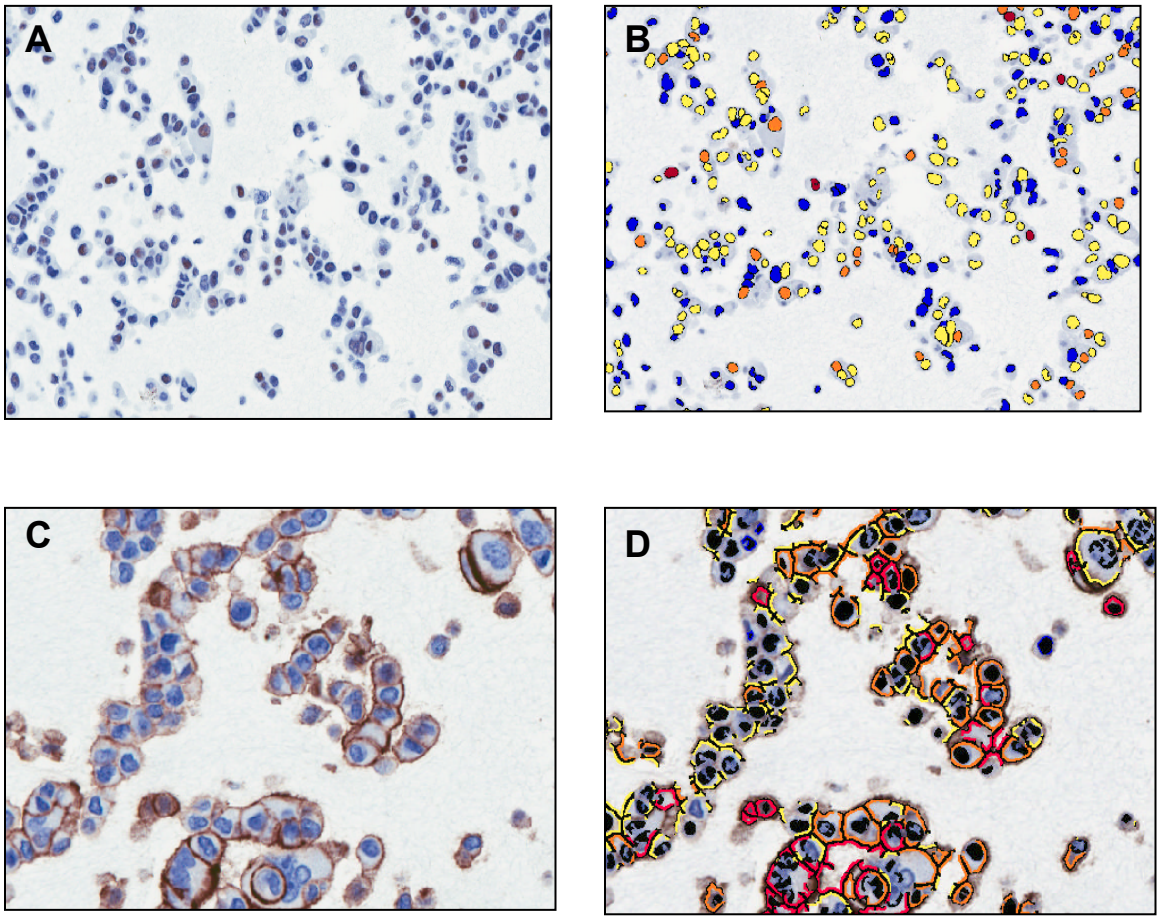
Slides prepared from FFPE blocks of nine human breast cell lines (AU-565, BT-474, HBL-100, HS-578T, MDA-MB-231, MDA-MB-435, MCF-7, T47D, SKBR-3) were processed in the same manner as human tissue samples at the SCH Laboratory. FFPE Cell line IHC-staining for ER, PR and HER2 were observed to be comparable to that of FFPE tissue samples. Figure 2 presents sample images of such nuclear and membranous IHC staining of FFPE tissue and cell lines. IHC staining of cell lines was confirmed by the identical IHC staining of non-paraffin embedded EZ-Slide and cytospin cell line preparations. All FFPE cell line slides were digitally scanned and analyzed with IA. Analysis overlay images confirmed that cell lines were accurately recognized and scored by the software (Figure 2). The resulting data was used to calculate HScores for project analysis.

3.1.1.1 IA reproducibility of IA and reproducibility of cell block samples (biological homogeneity)

A single slide section from the first created cell line block (T47D) was stained for PR at SCH laboratory and scanned 10 separate times. Each separate scan image was analyzed with IA between one and five separate times. The resulting calculated mean HScore for all scans and IA scoring was 282.1 ± 0.3 (Table 1). This demonstrated reproducibility of the Aperio IA system for scoring FFPE IHC-stained samples.

Next, the biological homogeneity of cell lines was examined throughout an entire FFPE cell line block. A single T47D cell line block was completely sectioned onto slides. Random sections from throughout the depth of the block were stained for PR at SCH Laboratory. PR IHC was selected because it tends to be less reproducible in human

Figure 2: IA Scoring and Calculation of HScore. Stain scores can be used to calculate an HScore which reflects the overall total staining of the tissue section. HScore= (3*(3+)) + (2*(2+)) + (1*(1+)). A, nuclear IHC staining (ER and PR); B, nuclear staining with IA analysis; C, membranous IHC staining (HER2); D, membranous staining with IA analysis.



0=blue, 1+= yellow, 2+= orange, 3+= red.

Table 1. Reproducibility of IA for IHC PR Staining.[a]

Scan	Algorithm Rep	(3+) % Nuclei	(2+) % Nuclei	(1+) % Nuclei	(0+) % Nuclei	HScore[b]
1	1	91.0	3.5	1.7	3.8	281.7
1	2	91.0	3.5	1.7	3.8	281.7
1	3	91.0	3.5	1.7	3.8	281.7
1	4	91.0	3.5	1.7	3.8	281.7
1	5	91.0	3.5	1.7	3.8	281.7
2	1	91.4	3.3	1.6	3.7	282.4
2	2	91.4	3.3	1.6	3.7	282.4
2	3	91.4	3.3	1.6	3.7	282.4
2	4	91.4	3.3	1.6	3.7	282.4
2	5	91.4	3.3	1.6	3.7	282.4
3	1	91.3	3.3	1.6	3.7	282.3
3	2	91.3	3.3	1.6	3.7	282.3
3	3	91.3	3.3	1.6	3.7	282.3
3	4	91.3	3.3	1.6	3.7	282.3
3	5	91.3	3.3	1.6	3.7	282.3
4	1	91.4	3.3	1.6	3.7	282.5
4	2	91.4	3.3	1.6	3.7	282.5
4	3	91.4	3.3	1.6	3.7	282.5
4	4	91.4	3.3	1.6	3.7	282.5
4	5	91.4	3.3	1.6	3.7	282.5
5	1	91.1	3.4	1.7	3.8	281.8
5	2	91.1	3.4	1.7	3.8	281.8
5	3	91.1	3.4	1.7	3.8	281.8
5	4	91.1	3.4	1.7	3.8	281.8
5	5	91.1	3.4	1.7	3.8	281.8
6	1	91.4	3.3	1.6	3.8	282.3
7	1	91.4	3.3	1.6	3.7	282.5
8	1	91.2	3.4	1.7	3.7	282.1
9	1	91.4	3.3	1.6	3.7	282.4
10	1	91.4	3.4	1.6	3.6	282.6
					Mean	**282.1**
					SD	**0.3**

[a]Multiple scans and analyses of one PR-stained T47D cell block slide at different times.

[b]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

tissue samples. IA of 20 of these different levels resulted in a mean HScore of 284.7 ± 2.4. This demonstrated biological homogeneity and consistency of staining for cell lines throughout the depth of the FFPE block (Table 2).

Cell lines were divided up into Low (HScore <50), Medium (HScore, 50-150) and High (HScore >150) groups. Although there was a variation in stain intensity of cell line blocks from run to run, cell-line HScores were reproducible and were not significantly different throughout the numerous IHC-stain runs. Stain results for the nine cell line blocks throughout all the different the study trials can be seen in Table 3. Slides from the initial FFPE cell line blocks and the Final CLMA block (discussed later) were stained at the SCH Laboratory. EQA IHC results for cell lines are an average of HScores from many laboratories.

3.1.1.2 Limitations of IA

When scanning tissue samples, a difficulty with IA was observed. IA algorithms could not be modified to reliably differentiate tumor tissue from normal tissue. Consequently, areas for analysis must be specifically selected or screened for IA prior to analysis to ensure that non-valid areas are not measured and erroneously used for evaluation. It was observed that damaged or folded TMA cores are erroneously scored by IA to have falsely higher IA intensity score values. Overlapping and damaged tissues have a darker colour and are therefore interpreted by IA colour measuring algorithms to be more intensely stained than they actually are. In Figure 3, circles denote areas of cores where IA-measured stain intensities were incorrectly measured. IA algorithms may also, rarely, incorrectly record a missing TMA core as a negatively staining sample. Consequently, all slides used in analysis in this project were individually inspected to remove IA error due to lost, folded or damaged samples.

3.1.2 IA in PT: cIQc Run 13 and cIQc Special Run

For IA of PT samples, sectioned TMA slides were stained and returned from participating cIQc Program laboratories. For Run 13, slides stained for ER, PR and HER2 were received from 46, 45 and 37 laboratories, respectively. Of these slides, 46, 41 and 34 laboratories were of sufficient quality to be analyzed by IA for ER, PR and HER2, respectively. Run 13 TMA cores were analyzed and scored as either positive or

Table 2. Cell-line Block IHC PR Staining Consistency.[a]

Slide	(3+) % Nuclei	(2+) % Nuclei	(1+) % Nuclei	(0+) % Nuclei	HScore[b]
1	92.0	3.0	1.7	3.4	283.6
2	93.3	2.5	1.3	3.0	286.0
3	89.2	3.9	2.7	4.2	278.2
4	93.6	2.1	1.3	3.0	286.2
5	93.9	1.7	1.1	3.3	286.1
6	93.7	2.1	1.2	3.0	286.5
7	93.5	2.4	1.3	2.9	286.4
8	92.8	2.6	1.4	3.2	285.1
9	94.5	1.8	1.0	2.7	288.1
10	92.5	2.9	1.5	3.1	284.9
11	92.8	2.9	1.3	3.0	285.6
12	93.8	2.3	1.1	2.8	287.1
13	93.7	1.9	1.3	3.0	286.4
14	93.3	2.7	1.3	2.7	286.7
15	91.5	3.1	1.8	3.6	282.5
16	91.5	3.3	1.9	3.3	283.0
17	92.2	3.1	1.8	2.9	284.7
18	92.1	3.1	1.6	3.2	284.2
19	90.9	3.5	2.0	3.6	281.7
20	90.7	3.9	2.0	3.4	282.0
				Mean	284.7
				SD	2.4

[a]Scanned and analyzed PR-stained IHC slides from random sections throughout the depth of one T47D cell bock.

[b]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

Table 3: Summary of FFPE Cell Line Block IHC HScores in Different IHC Runs[a]

ER

	Run 13 TMA[b]	Special CLMA[b]	FFPE Cell Block[c]	Final TMA[d]
Cell line / Number of slides	46	40	3-9	40
AU-565	54 ± 41	47 ± 47	3 ± 3	15 ± 14
BT-474	155 ± 56	167 ± 71	108 ± 19	58 ± 17
HBL-100	55 ± 46	30 ± 39	8 ± 1 0	11 ± 15
HS-578T	46 ± 36	39 ± 40	11 ± 9	7 ± 9
MCF-7	238 ± 38	250 ± 36	182 ± 40	187 ± 29
MDA-MB-231	61 ± 53	49 ± 46	3 ± 4	19 ± 18
MDA-MB-435	51 ± 46	45 ± 52	9 ± 11	17 ± 17
SKBR-3	64 ± 50	---	1 ± 1	14 ± 14
T47D	208 ± 57	224 ± 56	160 ± 42	160 ± 42

PR

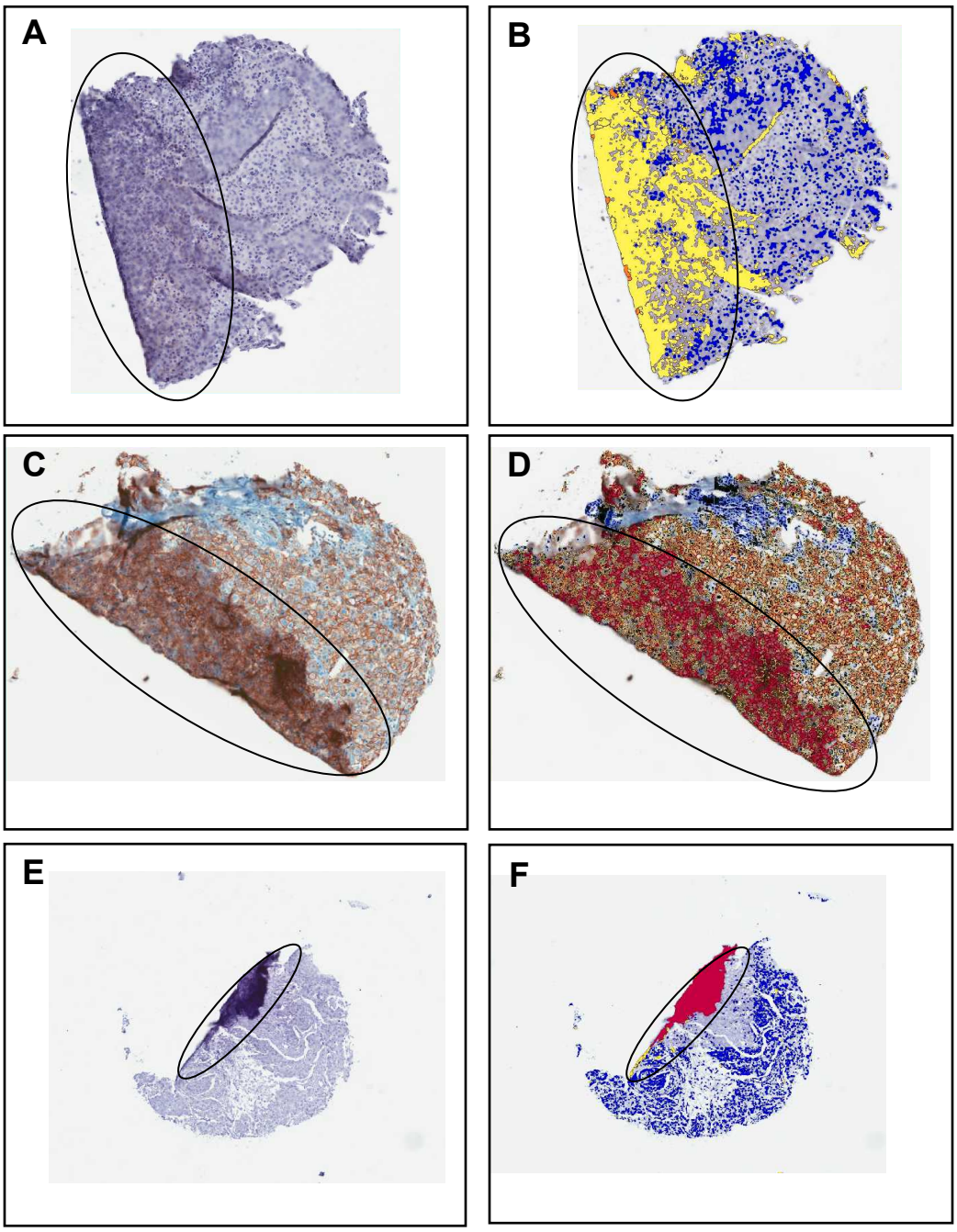
	Run 13 TMA[c]	Special CLMA[c]	FFPE Cell Block[d]	Final TMA[e]
Cell line / Number of slides	41	40	2-6	40
AU-565	53 ± 45	48 ± 60	7 ± 3	15 ± 14
BT-474	275 ± 16	213 ± 47	243 ± 19	237 ± 12
HBL-100	50 ± 42	34 ± 43	8 ± 7	21 ± 29
HS-578T	42 ± 33	36 ± 36	28 ± 10	16 ± 22
MCF-7	49 ± 43	41 ± 50	3 ± 2	24 ± 21
MDA-MB-231	42 ± 41	44 ± 50	15 ± 5	21 ± 21
MDA-MB-435	158 ± 41	142 ± 47	110 ± 41	143 ± 19
SKBR-3	64 ± 45	---	7 ± 5	14 ± 13
T47D	294 ± 6	287 ± 30	285 ± 6	282 ± 6

HER2

	Run 13 TMA[c]	Special CLMA[c]	FFPE Cell Block[d]	Final TMA[e]	JGH Control[f]
Cell line / Number of slides	34	35	4-9	40	54
AU-565	264 ± 73	247 ± 82	268 ± 13	263 ± 7	275 ±19
BT-474	279 ± 10	210 ± 94	265 ± 3	250 ± 9	---
HBL-100	86 ± 88	15 ± 22	46 ± 15	2 ± 2	---
HS-578T	79 ± 73	25 ± 22	7 ± 4	1 ± 1	---
MCF-7	177 ± 79	45 ± 52	66 ± 14	10 ± 8	---
MDA-MB-231	76 ± 88	17 ± 27	7 ± 5	1 ± 2	0.3 ± 0.4
MDA-MB-435	51 ± 77	21 ± 21	0 ± 0	0 ± 0	---
SKBR-3	294 ± 13	---	255 ± 26	269 ± 5	---
T47D	220 ± 67	97 ± 66	170 ± 17	142 ± 30	---

[a]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).
[b]Results expressed as average HScore ± SD.
[c]Run 13 and Special TMA, CLMA created for cIQc EQA runs, containing cell line and tissue cores.
[d]FFPE Cell blocks, the formalin-fixed paraffin-embedded cell line blocks stained at SCH Laboratory.
[e]Final TMA,CLMA created for in-laboratory PT at SCH Laboratory, containing cell line and tissue cores.
[f]JGH Controls, TMA control slides created at JGH for HER2 IHC.

Figure 3: Erroneous IA of Stained, Folded or Damaged TMA Cores. Circled areas denote where IA-measured stain intensities are incorrectly measured. The differences in colour between the undamaged portion of the core and the damages areas in IA analysis indicate differently measured stain intensities. IA images (B, D and F) use colour to indicate stain intensities (0, blue, 1+, yellow, 2+, orange, 3+, red). A, IA image of a folded nuclear IHC-stained core; B, incorrect IA analysis of image A; C, IA image of a folded membranous IHC-stained core; D, incorrect IA analysis of image C; E, IA image of a damaged nuclear IHC-stained core; F, incorrect IA analysis of image E.



negative by participating laboratories. This information was submitted together with the stained slides to the cIQc Program to assess IHC-interpretation skill as a part of the EQA evaluation. Slides stained by participating laboratories for ER, PR, and HER2 were visually assessed by cIQc pathologists and technologists in Toronto, Ontario in October 2010 (so-called "expert assessment"). Participating laboratories with acceptable staining (≥90% agreement with reference results and/or a $\kappa \geq 0.8$) received a passing grade from the cIQc Program.

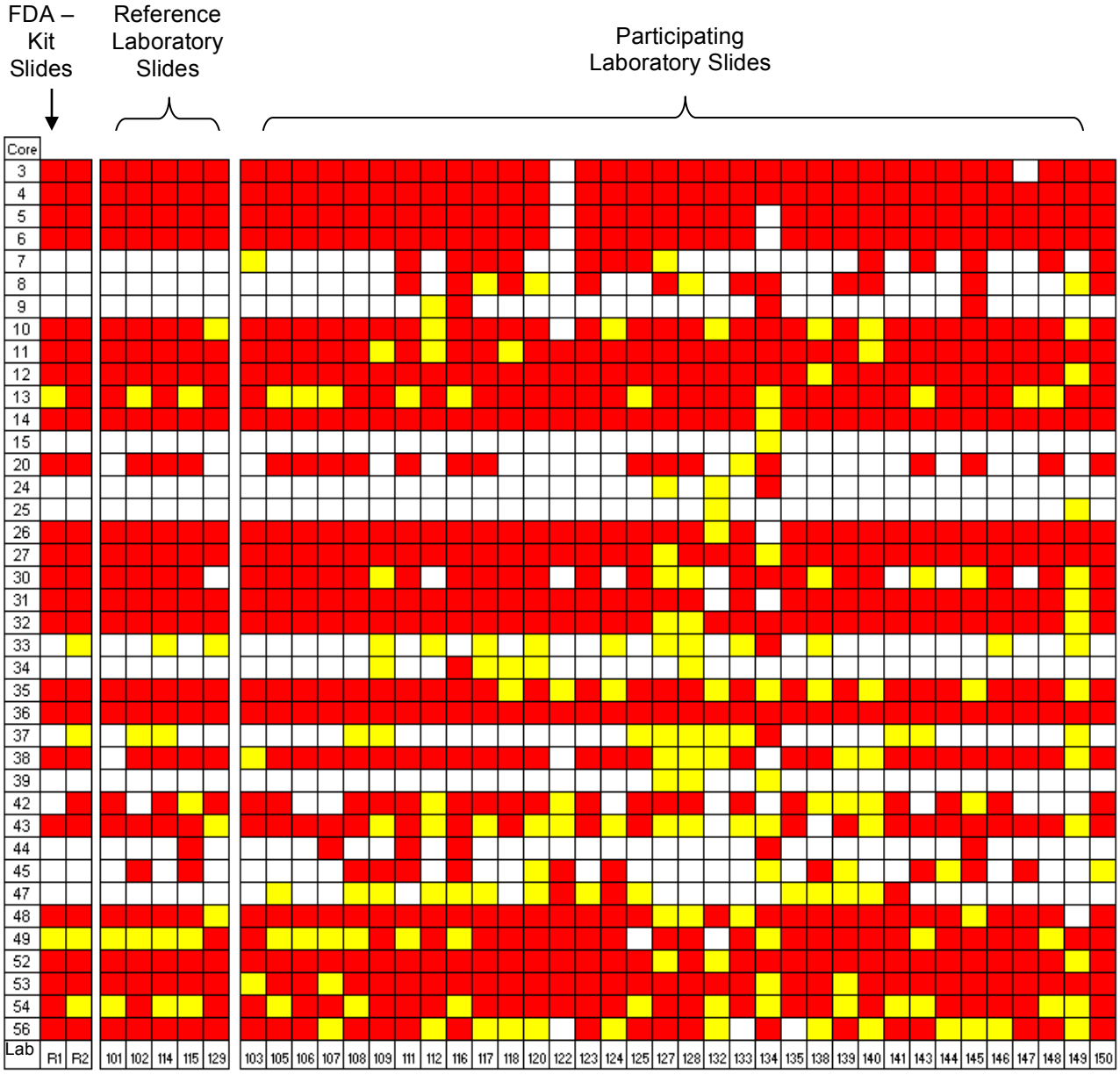
cIQc Program results are summarized using a garratogram. In a garratogram, tissue samples showing positive staining are denoted with red squares, negative results are denoted with white squares and unacceptable or missing samples are denoted as yellow squares. When the results of all cores for all participating laboratories are lined up beside those for reference laboratories, the cores that are not properly stained are easily observed. Figure 4 is an example of selected slides from Run 13 ER garratograms. Cell line samples were not included in the garratogram because their results were measured with IA for the purpose of this study and were not a part of the PT program.

The second “Special” cIQc run CLMA contained the newly created FFPE cell lines and FFPE breast carcinoma cores donated by JGH. Slides were returned from 40, 40 and 35 laboratories for ER, PR and HER2, respectively. Of these slides, 32 could be measured by IA for ER, 26 for PR and 32 for HER2. Unfortunately, the donated tissue cores were suboptimal and did not adhere well to the slides, resulting in unusable tissue IA results from this run. Consequently, only Run 13 results were available to assess how results obtained by participating laboratories with tissue cores compared with that of cell lines. The Special cIQc Run cell line cores results were utilized differently. IA HScores and LSRSRs of all cell lines were used to compare PT performance between two consecutive cIQc Program runs (Run 13 and Special Run).

3.1.2.1 Using IA data for PT performance evaluation: LSRSR and HScore

Slides from PT were stained by many different laboratories using differing methods. With no standardized IHC protocols or controls, a wide range in measured IHC results is often detected in most PT schemes/runs. The results of PT are usually evaluated

Figure 4: Garratogram of Expert Assessment Results from clQc ER Run 13. Garratograms allow for visual comparison of expert pathologist assessment of slides stained by an FDA-approved kit and by reference laboratories to expert pathologist assessment of EQA slides stained by participating laboratories. Laboratories considered successful at EQA garratogram evaluation have > 90% agreement with reference slides. Red squares, positive stain (>1% cell nuclei staining); White squares, negative staining; Yellow squares, Unacceptable/missing cores. Results are shown for 39 of the total 58 participating laboratories.



qualitatively and only partly quantitatively, since no measuring takes place. Consequently, the parameters of test performance (statistical analyses and comparisons of sensitivity and specificity) that are traditionally used in quality management are impractical and, most of the time, impossible. Specific laboratory information was unavailable to stratify EQA laboratories into sub-groups (by HIER method, antibody or reagents used) for further comparison. LSRSR and graphical methods were examined as useful methods to better quantify IHC PT results.

3.1.2.1.1 LSRSR analysis and calculation

In our 2011 publication, LSRSR was proposed as a new ratio for the monitoring of EQA IHC performance [161]. The closer an LSRSR is to a value of 1, the closer the performance of the laboratory is to that of the reference laboratory. Initial detailed statistical analysis of Run 13 HScore LSRSRs was carried out with one reference laboratory (cIQc Laboratory 114), which showed the most consistent performance, from one run to another, from inception of the cIQc Program. The LSRSR performance of laboratories was divided into ranges of Low (<0.8), medium (0.8-1.2) and high (>1.2).

Run 13 ER sensitivity and specificity were determined for single-reference LSRSR. LSRSR was compared to the garratogram scoring for each tissue core of each slide. Laboratory 114 results were selected by the cIQc Program as the true positive and negative values for both calculations. Calculation of LSRSRs for all cores on all slides was time consuming and labour intensive. By this approach, five laboratories performed unsatisfactorily (<80% sensitivity) for Run 13 ER tissue cores.

Although it was not statistically significant, it interesting to note that >91% (14 cores) of the poorly stained cores in Run 13 ER EQA (as rated by expert assessment) have single-reference LSRSRs outside of 0.8-1.2 (range, 0-5; average, 0.6). These initial calculations demonstrated how LSRSRs are complementary to expert assessment results, and can be used instead of expert assessment when the methods applied to generate LSRSRs are properly validated. Although analysis with garratograms appears simpler and less labour-intensive than determining single-reference LSRSRs, the effort and expertise required to generate garratograms is greater than the calculation of LSRSRs when a smaller number of samples is used for evaluation.

3.1.2.1.2 Using IA HScore to graphically demonstrate PT performance: boxplot and histograms

Box plots and histograms offer excellent simple visual summaries of EQA HScore results. cIQc Run 13 results were plotted for each core of each slide from all participating laboratories. cIQc results were compared to LSRSR results for Run 13 using box-plots and histograms. LSRSR graphs are very similar to HScore graphs when used in this manner. While LSRSRs are more efficient in showing variation with respect to a reference value, HScores are more convenient, time-wise. Consequently, HScores were used for these visual comparisons.

Histograms and boxplots visualize individual EQA laboratory performance with respect to the entire group. Laboratories with poor performance had HScores far from the mean of the results. A low or high LSRSR reflects this poor performance. It is possible to distinguish both poor and adequately performing laboratories on a histogram or boxplot of EQA results for TMA cores. Figure 5 shows an example of laboratories with a good and a poor LSRSR, using boxplots and histograms for an individual TMA core.

3.1.3 Optimal use of EQA LSRSR: multiple-reference evaluation of cIQc EQA runs

Even the best reference laboratory in the cIQc Program showed some variation in IHC results. The limitation of using single-reference LSRSR is in the fact that the samples used for testing are human tumors, which always show biological variation within a tissue core. Consequently, for this study, a different approach was used to partly compensate for this tissue heterogeneity. A small number of representative, diagnostic cores were evaluated from slides stained by a group of reference laboratories. A small number of tissue cores with informative HScores were selected, since the majority of high expressing and low/negative expressing samples were not contributory in our analyses. This multiple-reference approach was used for the remainder of this research. Once LSRSRs for all reference laboratories and cores were plotted, the minimum and maximum acceptable LSRSRs were identified. Example graphs of multiple-reference LSRSR acceptable ranges for cell line and tissue cores in Run 13 ER EQA are shown in Figure 6.

Figure 5: Box Plot and Histogram Plots of HScores of One Tissue Core for a Good and a Poor Performing Laboratory. Box plots and histograms were used to summarize EQA HScore results from cIQc Run 13 of ER for all participating laboratories. HScores for a good and a poor performing laboratory are indicated by arrows in relation to the group results. A, a good performing laboratory; B, a poor performing laboratory.

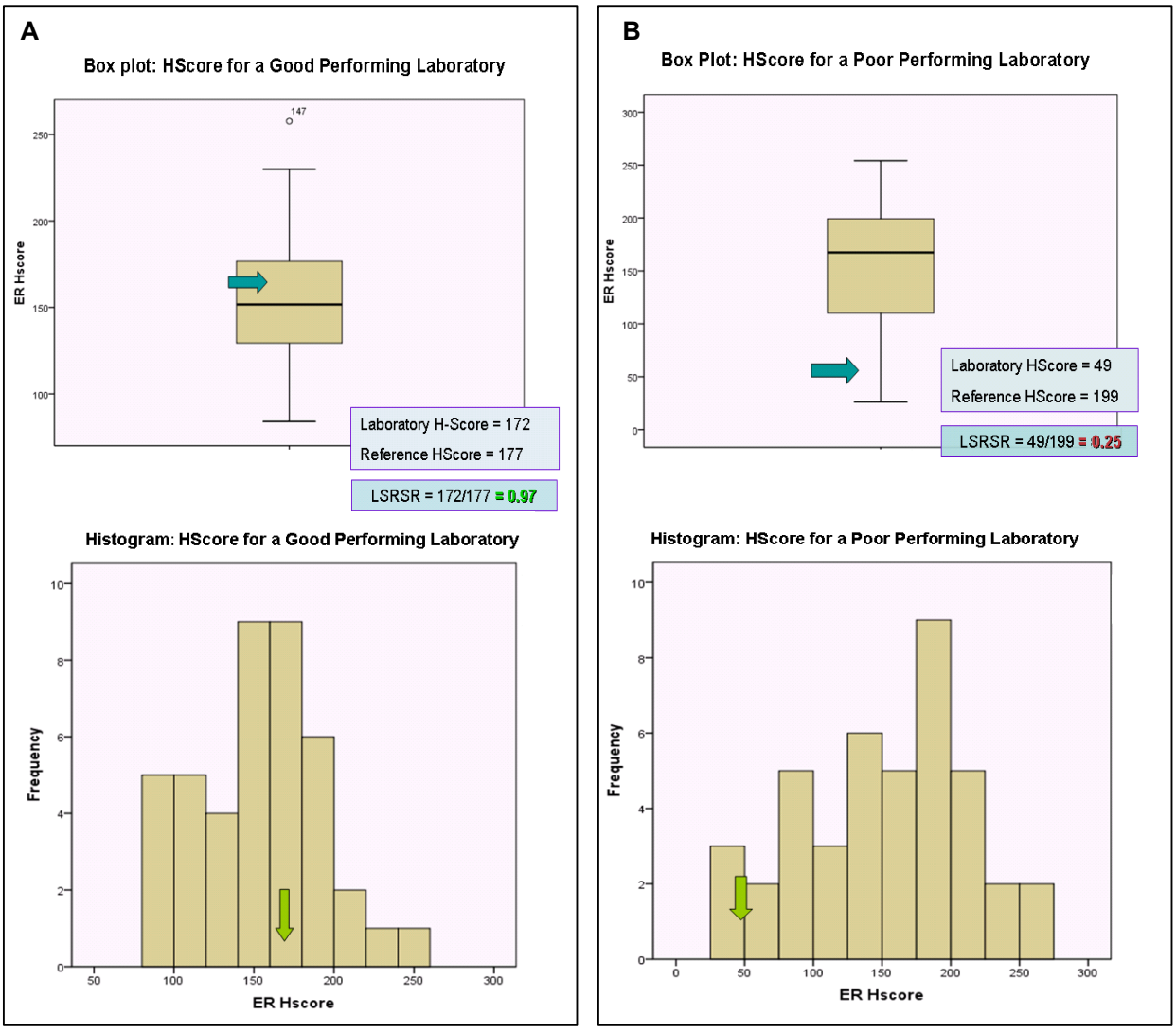
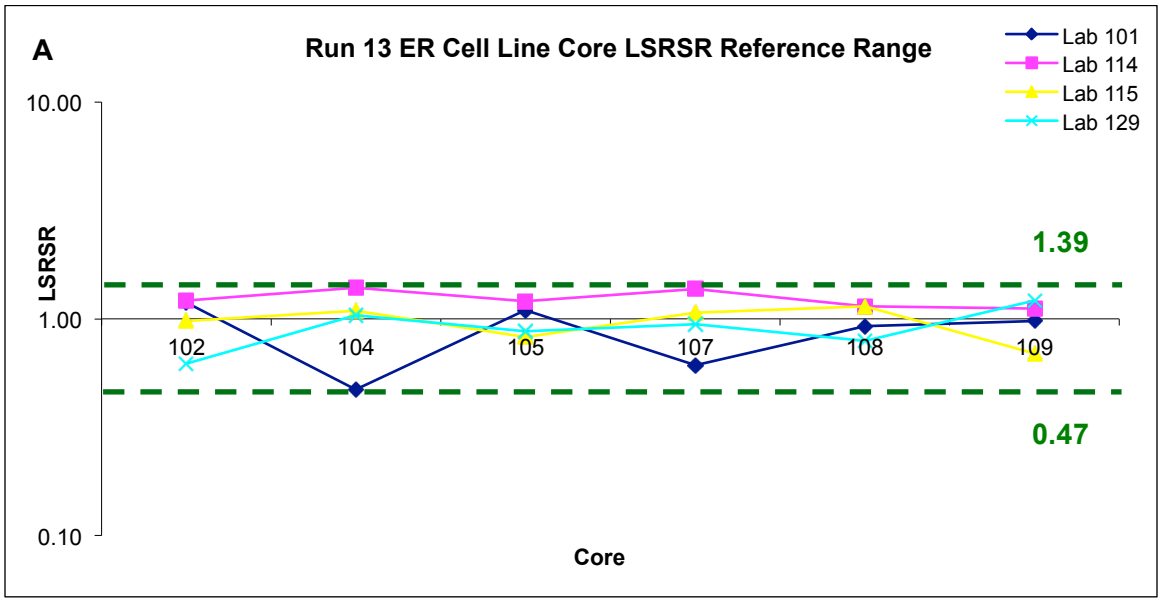
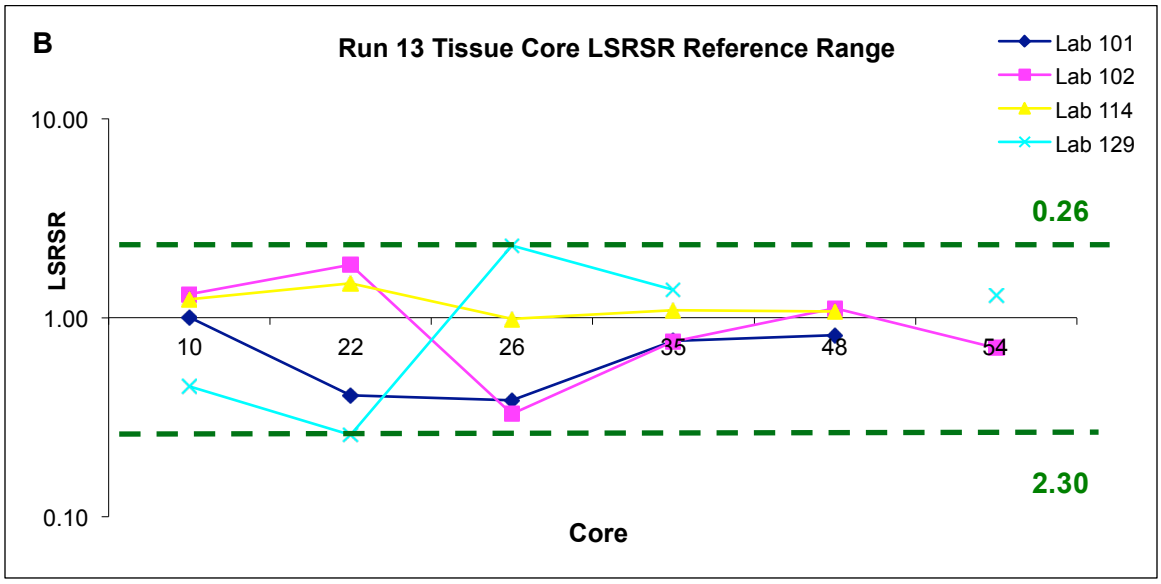


Figure 6: **LSRSR Range Assessment for Cell Line and Tissue Cores in Run 13 ER EQA**. Six representative cores were selected and the HScores from five reference laboratories were used to calculate LSRSRs for these cores. The minimum and maximum of these reference laboratory LSRSRs are the acceptable LSRSR range for the EQA performance rating of all participating laboratories in Run 13 ER. A, LSRSR range for cell line reference cores; B, LSRSR range for tissue reference cores.





Representative cores with mean HScores in the linear diagnostic IHC range were selected for Run 13 and Special Run ER, PR and HER2. Six reference laboratories (101, 102, 114, 115, 126, 129) were selected based on consistent previous EQA cell line cores and acceptable LSRSR ranges defined for each IHC stain and core type (cell line or tissue).

3.1.3.1 LSRSR reference ranges for cIQc EQA runs

3.1.3.1.1 Run 13 LSRSR assessment of PT performance

For ER and PR, six cell line and tissue cores were selected by mean HScore. For HER2, five Run 13 cell line and tissue cores were selected by mean HScore. HER2 cell line core selection was problematic because most FFPE cell lines stained very low (HScore <20) or high (HScore >150). Consequently, two HER2 cores with HScores above 100 were selected for LSRSR calculations. Cell lines AU-565, HBL-100, HS-578T, MCF-7, MDA-MB-435 and MDA-MB-231 were used for evaluation of ER (mean HScores, 60-98). Cell lines AU-565, BT-474, HBL-100, HS-578T, MDA-MB-231, MDA-MB-435 and SKBR-3 were used for evaluation of PR (mean HScores, 43-73). Cell lines AU-565, HS-578T, MCF-7, MDA-MB-435, and T47D were used for HER2 evaluation (mean HScores, 50 - 259). The selected ER, PR and HER2 tissue cores had mean HScores of 51- 93, 41 - 105 and 35 – 103, respectively.

In each analysis, one or more reference laboratories were excluded. One laboratory did not participate in Run 13. Other laboratories were excluded from ER, PR and HER2 analyses due to extreme LSRSR values (as per Section 2.6.3.4). The number of reference laboratories, representative cores and acceptable reference LSRSR ranges for Run 13 ER, PR and HER2 cell line and tissue cores can be found in Table 4.

Initially, LSRSR was applied as a “pass/fail” measurement, with acceptable laboratory performance defined as having >80% of the selected array cores with acceptable LSRSR values. For ER and PR, cell line LSRSR ranges for reference laboratories were smaller and more discriminatory, resulting in a lower percentage of acceptably performing laboratories. In contrast, for HER2, tissue LSRSR ranges were

Table 4: Run 13 Cell Line Core and Tissue Core LSRSR[a] Reference Ranges[b] Used in EQA Analysis of Performance. A, LSRSR reference calculation data. B, IA LSRSR evaluation of overall laboratory performance as a pass/fail measurement.

A.

	Cell Line Cores			Tissue Cores		
	ER	PR	HER2	ER	PR	HER2
Reference laboratories[c]	4	3	4	4	4	4
Representative cores[d]	6	6	5	6	6	5
IA LSRSR reference ranges	0.47-1.39 [e]	0.35-1.67 [e]	0.27-2.05 [f]	0.49-1.72 [e]	0.24-2.05 [e]	0.33-1.73 [f]

B.

Value	ER		PR		HER2	
	Tissue	Cell	Tissue	Cell	Tissue	Cell
Slides received from laboratories	46	46	45	45	37	37
TMA sample cores measureable by IA[g]	42	44	52	48	34	34
Laboratories measured with IA	39	46	41	41	34	33
LSRSR acceptable laboratories[h]	17	16	28	19	23	24
% LSRSR acceptable laboratories[h]	44%	35%	68%	46%	68%	73%
Laboratories acceptable by both cell and tissue LSRSR[h]	12		19		16	
Laboratories unacceptable by both cell and tissue LSRSR[h]	12		6		3	

[a]LSRSR; Laboratory Score: Reference Method Score Ratio, calculated from IA HScore values.

[b]LSRSR reference ranges are calculated using the average HScore values of the designated reference laboratories for five or six representative cores.

[c]Reference Laboratories are selected by cIQc for each run based on previous high quality of IHC staining. Not all laboratories could be evaluated for all stains due to loss of tissue in representative cores.

[d]Representative cores are selected with HScores between 40 and 100 so that staining levels are in an optimal linear range.

[e]Cell line LSRSR ranges are more stringent than tissue LSRSR ranges due to tissue core heterogeneity. More laboratories were acceptably rated when the less stringent tissue LSRSR was used.

[f]HER2 LSRSR ranges are the opposite in stringency to those for ER and PR: Tissue LSRSR are more stringent. Lack of ideal diagnostic HER2 reference cores may be the cause.

[g] Some core loss occurred. Consequently, a number of laboratories were evaluated with as many of the six representative cores that remained on the slide. LSRSR values were calculated from no less than three and no more than five reference laboratory HScores.

[h]Acceptable laboratories had ≥ 80% of their calculated LSRSR values fall with the reference LSRSR ranges for the representative cores. Passing criteria is >80% of the six representative cores having an LSRSR within the LSRSR reference range.

more stringent, with a lower percentage of laboratories given an acceptable performance evaluation. Tissues or cell lines with high HScores were less sensitive for detecting variation of IHC staining. Cores with very low HScores were more variable in LSRSR analysis due to the larger relative differences in the small values used for LSRSR calculation. Not all laboratories with acceptable performance were rated as such by both cell line and tissue LSRSR. Pass/fail results are summarized in Table 4.

3.1.3.1.2 Run 13 cIQc assessment of PT performance

cIQc Program EQA evaluation of Run 13 was presented using two methods: the calculation of agreement of tissue core staining with reference staining and a comparison of laboratory stain results to reference stain results using κ values. Calculated agreement with a reference (and evaluation by κ) concluded that 60% (69%), 75% (73%) and 95% (93%) of laboratories had acceptable performance (as defined by international standards for these particular tests) for ER, PR and HER2 IHC, respectively. Twelve, three and two laboratories had sub-optimal performance (80-90% concordance with reference) for ER, PR and HER2, respectively. The cIQc Program notified laboratories of any unacceptable performance. Suggestions for improvement to submitted laboratory IHC protocols were given, whenever possible. Detailed cIQc Run 13 assessment results are in Supplemental Table 2.

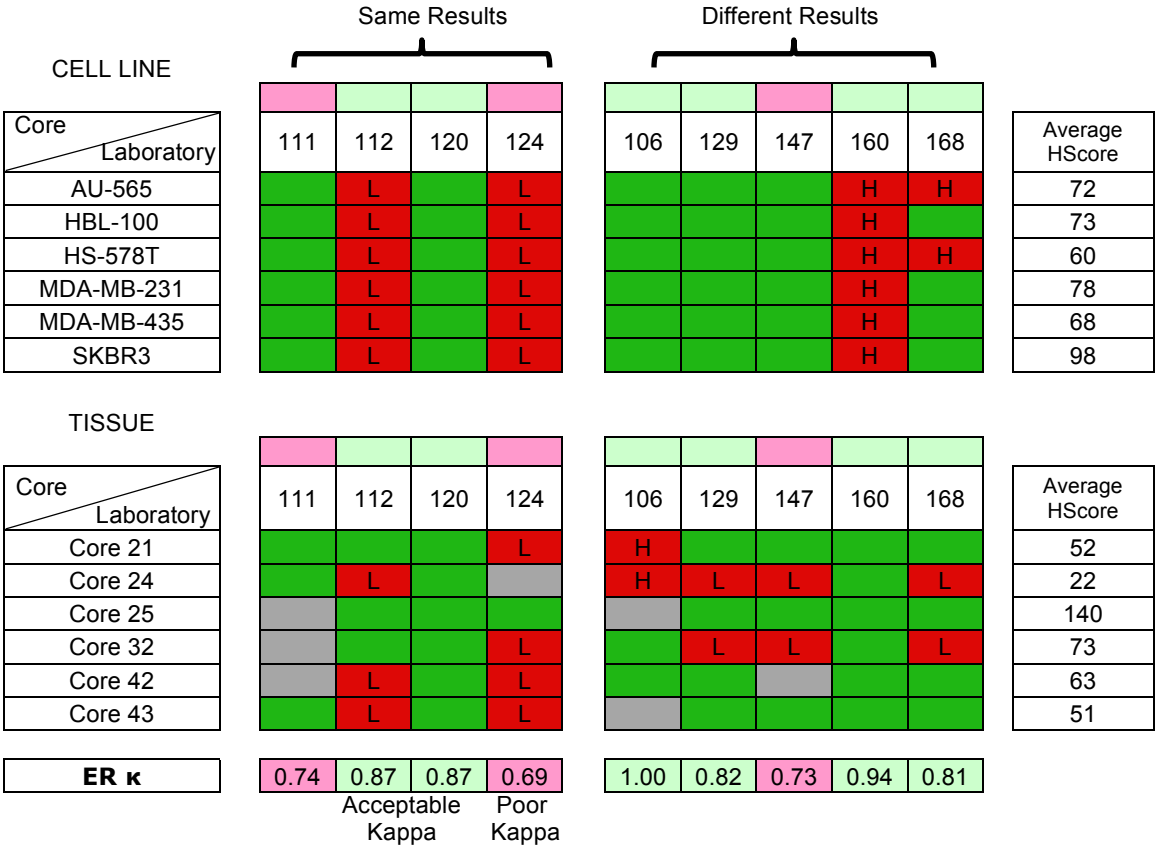
3.1.3.1.3 Comparison of Run 13 cIQc and LSRSR evaluation

Run 13 LSRSR results were compared to cIQc EQA assessment results. A figure was assembled, similar to a garratogram, to visually compare cIQc and LSRSR performance data. In Figure 7, examples of laboratories with agreeing cIQc and LSRSR results (both acceptable and unacceptable) and examples where the cIQc and LSRSR results disagree are shown. Table 5 then summarizes the overall agreement in ER, PR and HER2 staining by the two methods. LSRSR reference ranges for ER and PR cell lines appear to be more discriminatory than cIQc evaluation. In contrast, LSRSR reference ranges for HER2 tissues were more discriminatory than cell lines, and more discriminatory than cIQc expert assessment evaluation.

Cell line and tissue cores were examined to see whether the same types of unacceptable errors in staining were observed for a laboratory using LSRSR evaluation

Figure 7: A Visual Comparison of Run 13 ER Cell Line and Tissue Core IA LSRSR[a] and cIQc Assessment[b] Results. Selected examples of each type of agreement and disagreement are shown.

The top row of the figure indicates cIQc laboratory κ results; Light green, acceptable cIQc laboratory performance[c]; Light red, unacceptable cIQc performance. Remaining rows include IA LSRSR laboratory results for each cell line, with each laboratory as a separate column; Dark green, acceptable LSRSR performance[d]; Dark red squares, unacceptable LSRSR performance, with "H" and "L" unacceptable LSRSR values above and below acceptable the LSRSR range, respectively. Grey squares were inferior quality cores that could not be measured by IA.



[a]LSRSR, Laboratory Score: Reference Method Score Ratio, calculated from IA HScore values.

[b]cIQc assessment calculated from comparison of participant and cIQc pathologist-scored garratogram results

[c]Acceptable cIQc performance; 90% correct staining results and Cohen's κ ≥0.8 between laboratory IHC results and cIQc IHC-assessed staining results. Poor or unacceptable κ results were <0.8.

[d]Acceptable IA LSRSR performance; LSRSR of laboratory is within acceptable reference LSRSR range. LSRSR reference ranges are calculated using the average HScore values of the designated reference laboratories for five to six representative cores.

Table 5: Comparison of Acceptable[a] and Unacceptable[b] Run 13 EQA Results as Measured by IA LSRSR[c] and cIQc Assessment[d].

	ER		PR		HER2	
cIQc / LSRSR	**Cell Line**	**Tissue**	**Cell Line**	**Tissue**	**Cell line**	**Tissue**
N[e]	46	39	41	41	33	34
Acceptable cIQc / Acceptable LSRSR	10 (22%)	10 (26%)	17 (41%)	22 (61%)	21 (64%)	22 (65%)
Acceptable cIQc/ Unacceptable LSRSR	24 (52%)	18 (46%)	15 (37%)	7 (17%)	9 (27%)	9 (26%)
Unacceptable cIQc/ Acceptable LSRSR	6 (13%)	7 (18%)	2 (5%)	3 (7%)	3 (9%)	1 (3%)
Unacceptable cIQc / Unacceptable LSRSR	6 (13%)	4 (10%)	7 (17%)	6 (15%)	0 (0%)	2 (6%)

[a]Acceptable cIQc performance ≥ 90% correct staining results and κ ≥0.8 when compared to cIQc assessment garratogram values.

[b]Acceptable IA LSRSR performance, IA LSRSR of laboratory has 80% of cores within acceptable reference LSRSR range. LSRSR reference ranges are calculated from the average HScore values of the designated reference laboratories for five to six representative cores.

[c]LSRSR; Laboratory Score: Reference Method Score Ratio, calculated from IA HScore values.

[d]cIQc assessment calculated from comparison of participant and cIQc pathologist scored garratogram results

[e]N; Number of laboratories.

for both core types. Generally, there was an agreement between the cell line and tissue core evaluations for each laboratory. For tissue cores, 20 of 21, eight of eight and two of ten unacceptably performing laboratories had LSRSRs below the acceptable reference range, for ER, PR and HER2, respectively. For cell line cores, 23 of 30, 18 of 22 and five of nine unacceptably performing laboratories were below acceptable range for ER, PR and HER2, respectively. There were twelve, six and zero laboratories that performed unacceptably, in the same manner (either both high or both low), for both cell line and tissue LSRSRs for ER, PR and HER2, respectively. In all cases, the unacceptably performing ER and PR results were below acceptable range (false negative), indicating problems with test sensitivity. The six laboratories with unacceptable PR LSRSR results had unacceptable PR results by cIQc κ evaluation.

There were only three laboratories with disagreement in the type of unacceptable performance for their reference cores (i.e., staining both above and below the acceptable LSRSR ranges), suggesting problems with both sensitivity and specificity. For ER and HER2, one laboratory showed disagreement by having an unacceptably high cell line LSRSR and unacceptably low tissue LSRSR. For PR, one laboratory had one tissue core above and one tissue core below the acceptable LSRSR range. Examination of individual slides clarified these LSRSR disagreements as non-critical. The questionable ER and HER2 tissue cores were positively stained overall and in agreement with cIQc reference and κ scoring. Upon examination, the over-staining PR core did not contain any tumor tissue and was therefore excluded from analysis.

When an LSRSR was unacceptably high or low for laboratories rated acceptable by cIQc κ, the stained cores were visually examined and confirmed as positively stained for ER and PR, despite poor LSRSR results. This, together with the above results, demonstrates the use of LSRSRs as a calibration and monitoring tool.

3.1.3.2 Special Run LSRSR assessment of cell line EQA performance

Acceptable LSRSR ranges were calculated for Special EQA Run cell line cores to compare performance of participating laboratories over two consecutive EQA runs. Due to tissue loss, there were no sufficient tissue cores for cIQc expert assessment for this run. LSRSR reference ranges were defined using cell lines AU-565, HBL-100, HS-578T

MDA-MB-231 and MDA-MB-435 for ER; cell lines AU-565, BT-474, HBL-100, MDA-MB-231, and MDA-MB-435 for PR; and cell lines BT-474, HS-578T, MCF-7, MDA-MB-231, and T47D for HER2. Results from four reference laboratories were available. The acceptable LSRSR ranges were 0.69-1.50, 0.28-1.84 and 0.21-2.05 for ER, PR and HER2, respectively. There were nine, eleven and ten of 39 total participating laboratories with acceptable performance by LSRSR for ER, PR and HER2, respectively.

3.1.3.2.1 Run 13 and Special Run LSRSR EQA assessment comparison

Cell line LSRSRs for Run 13 and Special Run were compared to each other. LSRSR analysis identified subtle changes in IHC-staining results as a result of higher IA resolution and cell line biological homogeneity. Cell lines present in both runs with similar HScores were used for comparison. Laboratories with unacceptable performance, by cIQc κ assessment, in Run 13, were examined to determine whether a change/improvement of performance was observed with LSRSR in the next Special cIQc Run. Summary figures were assembled for comparison of ER, PR and HER2 results in the same manner as for Run 13 LSRSR and cIQc κ results. Figure 8 is a composite figure demonstrating this comparison, with examples of laboratories with no shift in LSRSR and ones with LSRSR shift in the subsequent run.

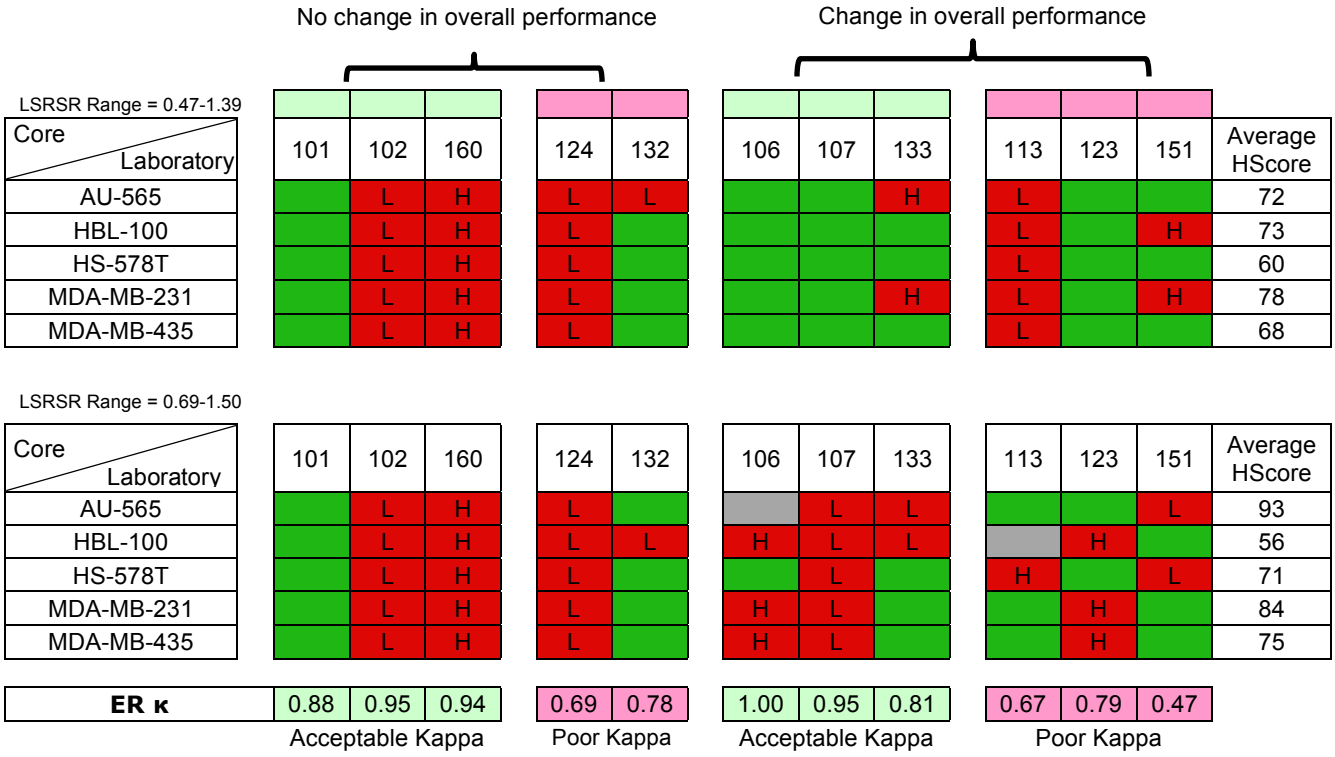
The ER reference samples compared between Run 13 and Special Run were cell lines AU-565, HBL-100, HS-578T, MDA-MB-231 and MDA-MB-435, with HScore Ranges of 60-78 and 56-93, respectively. The PR reference samples compared between Run 13 and Special Run were cell lines AU-565, HBL-100, MDA-MB-231, and MDA-MB-435, with HScore Ranges of 71-76 and 31-50, respectively. The HER2 reference samples compared between Run 13 and Special Run were cell lines BT-474 and T47D, with HScore Ranges of 259-253 and 58-122, respectively. There were 32, 26 and 26 laboratories with IA scorable results for both EQA runs for ER, PR and HER2, respectively.

3.1.3.2.2 Observed change in performance between subsequent runs (“Run 13” to “Special Run”)

Over 50% of laboratories who were informed about their unacceptable cIQc κ

Figure 8: Using IA LSRSR[a] to Observe Performance Change Between EQA Runs: A Comparison of Run 13 and Special Run ER Cell Line Evaluation Results. Summary of the laboratories with an unacceptable cIQc evaluation[b] in Run 13 which showed a measurable change in IHC-stain performance by IA LSRSR for the next cIQc Special Run. Selected examples of each type of agreement and disagreement are shown.

The top row of the figure indicates cIQc laboratory κ results; Light green, acceptable cIQc laboratory performance[c]; Light red, unacceptable cIQc performance. Remaining rows include IA LSRSR laboratory results for each cell line, with each laboratory as a separate column; Dark green, acceptable LSRSR performance[d]; Dark red squares, unacceptable LSRSR performance, with "H" and "L" unacceptable LSRSR values above and below acceptable the LSRSR range, respectively. Grey squares were inferior quality cores that could not be measured by IA.



[a]LSRSR, Laboratory Score: Reference Method Score Ratio, calculated from IA HScore values.

[b]cIQc assessment calculated from comparison of participant and cIQc pathologist scored garratogram results

[c]Acceptable cIQc performance; 90% correct staining results and Cohen's κ ≥0.8 between laboratory IHC results and cIQc IHC assessed staining results. Poor or unacceptable κ results were <0.8.

[d]Acceptable IA LSRSR performance; LSRSR of laboratory is within acceptable reference LSRSR range. LSRSR reference ranges are calculated using the average HScore values of the designated reference laboratories for five to six representative cores.

performance in Run 13 changed their performance in the Special Run for ER, PR and HER2. These were fully detectable by LSRSR while visual inspection was insensitive and uninformative. Table 6 summarizes the change in LSRSR measured stain performance between the Run 13 and Special EQA Runs. For ER, the LSRSR-measured change in stain performance for κ unacceptably performing laboratories was statistically significant (by Chi-squared evaluation of change in response) when compared to changes observed in κ acceptably performing laboratories. This demonstrates the use of LSRSR to measure IHC performance and method correction. There were too few HER2 unacceptable laboratories for statistical inference.

3.2 Daily positive controls monitored by IA through Levey Jennings QC charts.

Currently used, tissue-based, on-slide breast marker IHC controls for ER, PR and HER2 QA run from four Canadian hospitals were evaluated with IA. IA HScores of consecutive slides were plotted to create Levey-Jennings QC charts, as previously described. CLMAs containing FFPE tissue and cell lines were also created and stained for ER, PR and HER2 at SCH and for HER2 at JGH.

3.2.1 IA of currently used, tissue-based controls in clinical IHC

The functionality of the different tissue types used in the above on-slide IHC controls was compared. Consecutive, IA-measured and plotted sets of control slides came from the same control blocks. Breast tissue and/or endometrium tissues were used for ER and PR controls. Breast tissue was used for HER2 controls. IA HScore Levey-Jennings QC charts were examined to comment on best practise for measurement and application. Statistically significant differences in SD from the mean between whole and partial control IA HScore SD were measured with an F-test.

Trending in HScore could be observed over time, demonstrating IA HScore QC chart functionality. QC Charts were used for multiple comparisons of these current on-slide clinical tissue controls.

3.2.1.1 Effect of the area of IA analysis: whole *versus* partial analysis

On-slide tissue controls for ER, PR and HER2 were examined to see if IA of an entire control was necessary for QA purposes. The entire control and a smaller

Table 6: Using LSRSR[a] to Calibrate EQA IHC Performance: Change in IHC-Stain Performance in Special Run After an Unacceptable cIQc Run 13 Performance[b]. Special Run results were examined for a change in LSRSR[c] measured performance from Run 13. Laboratories with a poor Run 13 performance made more changes in their staining, as measured by changing IA LSRSR results in the next EQA Special Run.

	Acceptable (κ ≥ 0.8)		Unacceptable (κ <0.8)	
Stain	LSRSR Improvement	No change	LSRSR Improvement	No change
ER	6	18	4	4
PR	8	11	4	3
HER2	10	13	2	0

% Change LSRSR Results

Stain	κ ≥ 0.8	κ <0.8
ER	25%	50% *
PR	42%	57%
HER2	43%	100%

[a]LSRSR; Laboratory Score: Reference Method Score Ratio, calculated from IA HScore values.

[b]Acceptable Run 13 EQA performance is rated as a κ ≥ 0.8 between laboratory and cIQc IHC results.

[c]LSRSR reference ranges are calculated using the average HScore values of the designated reference laboratories for five to six representative cores.

*Changes in performance between the two EQA runs are statistically significant in cIQc unacceptably performing laboratories when compared to LSRSR changes occurring in cIQc acceptably performing laboratories. This suggests a calibration in methodology between runs.

representative portion of the same control were analyzed with IA from the same scanned image. QC charts for whole tissue and partial tissue IA HScore were compared. Whole IA HScore QC charts were observed to be more stringent, with smaller SD, for all locations and tissue types. The calculated difference in SD between whole and partial control was significantly different for many of the tissue controls for ER, PR and HER2 (Table 7, 8 and 9, respectively). ER controls had the largest number of significantly different controls. The QC chart HScore SD of all the partial control samples were larger than those of the whole control samples, regardless of the size of the smaller portion analyzed. An example of a control from JGH with corresponding QC charts, showing the difference between whole and partial section results, can be seen in Figure 9.

3.2.1.2 Effect of control tissue type

Breast and endometrium tissues were used separately, and in combination, as ER and PR IHC controls. Breast tissue was used as an HER2 IHC control. IA results for each type of on-slide control tissue were compared. Statistically significant differences in SD between whole and partial sample IA HScore SD were measured with an F-test. Significant differences were observed in SD for each tissue type in Tables 7, 8 and 9, for ER, PR and HER2, respectively. Although Ontario endometrium controls had no significant SD difference between whole and partial QC charts, a larger sample set is required to comment more definitively. No tissue was identified as a superior control.

3.3 Building better positive controls: comparison of new cell line-based controls to current tissue-based controls

A cIQc panel of pathologists visually selected low to intermediate staining FFPE cell lines from Run 13 cell line slides stained by reference laboratories. For ER, AU-565, HS-578T and MDA-MB-231 cells were selected, with MCF-7 cells selected as a high staining positive control. For PR, AU-565 and MCF-7 cells were selected, with T47D cells selected as a high staining positive control. For HER2, MCF-7 and MDA-MB-231 cells were selected, with AU-565 cells selected as a high staining positive control. For ER, PR and HER2, three, two and two cell lines selected by pathologists were also selected for use with IA, respectively.

Table 7. Description and IA Analysis of ER On-Slide IHC Controls.

ER Group	N[a]	Whole or Part	Mean HScore[b] ± SD	F-test p-value
Saskatoon (breast)	8	Whole Part	66 ± 14 158 ± 22	<0.01
	9	Whole Part	67 ± 12 134 ± 35	<0.01
	39	Whole Part	40 ± 42 77 ± 71	<0.01
	55	Whole Part	104 ± 33 123 ± 68	<0.01
	15	Whole Part	57 ± 36 102 ± 67	<0.01
	6	Whole Part	78 ± 16 121 ± 8	0.36
Ontario (endometrium)	11	Whole Part	192 ± 15 195 ± 14	0.6
Quebec (endometrium and breast)	30	Whole Part	89 ± 12 151 ± 27	<0.01
	15	Whole Part	82 ± 10 134 ± 17	<0.01
Alberta (breast)	71	Whole Part	76 ± 11 218 ± 13	<0.01

[a]N, number of slides for that unique control.

[b]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

Table 8. Description and IA Analysis of PR On-Slide IHC Controls.

PR Group	N[a]	Area	Mean HScore[b] ± SD	F-test p-value
Saskatoon (breast)	11	Whole Part	105 ± 11 215 ± 18	<0.01
	7	Whole Part	223 ± 12 219 ± 5	0.31
	7	Whole Part	219 ± 5 208 ± 42	0.06
	13	Whole Part	232 ± 5 113 ± 19	<0.01
	29	Whole Part	189 ± 13 114 ± 19	0.14
	43	Whole Part	228 ± 11 130 ± 28	0.01
Ontario (endometrium)	9	Whole Part	187 ± 17 187 ± 18	0.96
Quebec (endometrium and breast)	27	Whole Part	90 ± 12 64 ± 17	0.39
	15	Whole Part	87 ± 25 38 ± 17	0.78

[a]N, number of slides for that unique control.

[b]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

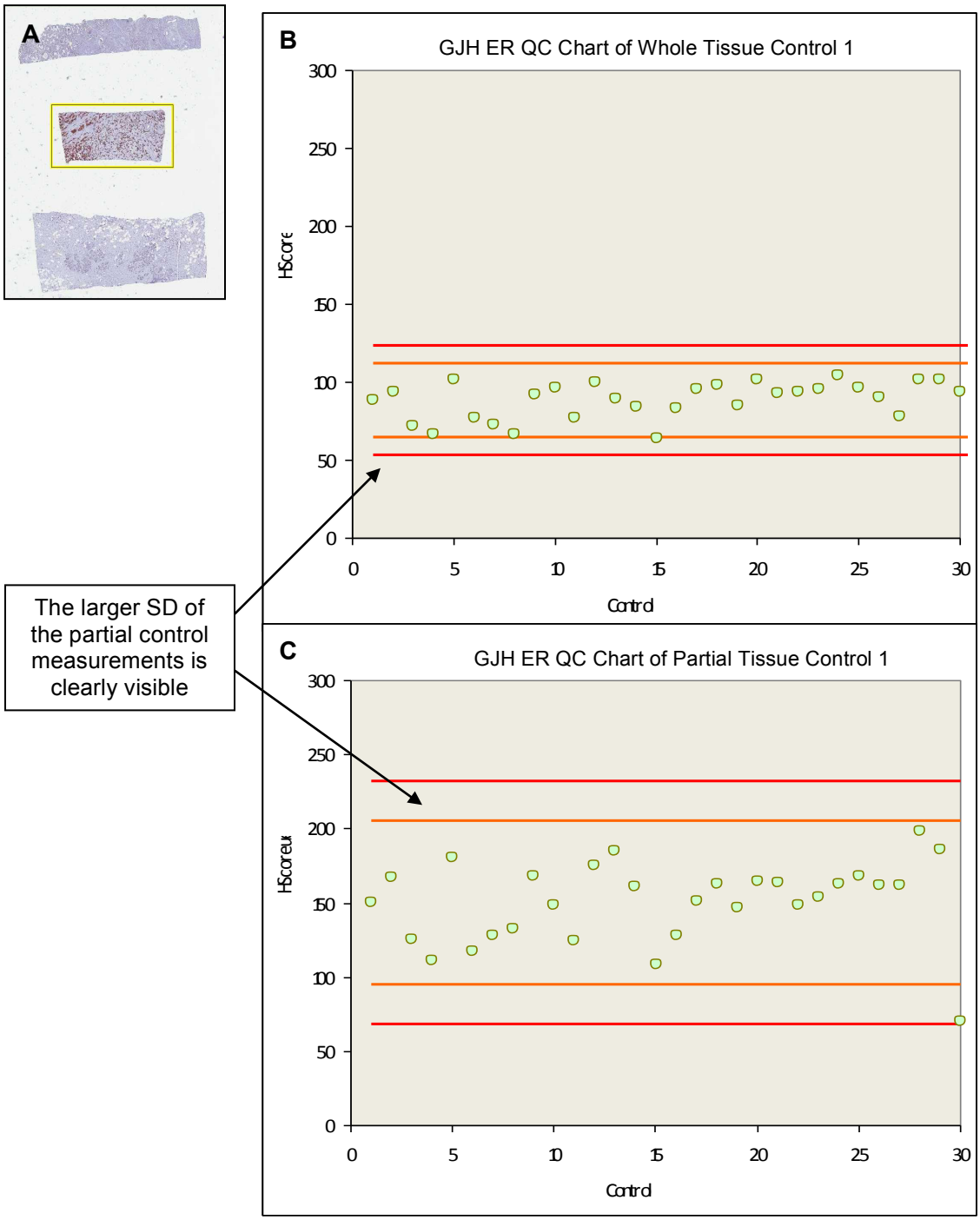
Table 9. Description and IA Analysis of HER2 On-Slide IHC Controls.

HER2 Group	N[a]	Area	Mean HScore[b] ± SD	F-test p-value
Saskatoon (breast)	18	Whole Part	104 ± 15 168 ± 56	<0.01
	89	Whole Part	112 ± 15 144 ± 17	0.35
	18	Whole Part	120 ± 24 148 ± 20	0.35
Quebec (breast)	58	Whole Part	73 ± 22 66 ± 20	0.14
Alberta (breast)	26	Whole Part	121 ± 22 177 ± 28	0.39
	25	Whole Part	75 ± 19 23 ± 10	0.01
	10	Whole Part	54 ± 9 15 ± 9	0.9
	15	Whole Part	128 ± 14 83 ± 25	<0.01

[a]N, number of slides for that unique control.

[b]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

Figure 9: A Comparison of Whole Section and Partial Section QC Charts. IA HScore values of thirty consecutive IHC-stained ER controls were plotted to create a QC chart. Orange lines denote two SD from the mean HScore. Red lines denote three SD from the mean HScore. Values falling outside of 2 SD range indicate a need for recalibration of method and/or instrument. Values outside of 3 SD indicate the need for repeated testing because such results are unacceptable. A, a stained section of the mixed tissue ER control, composed of breast and endometrium tissue; B, the whole control IA HScore QC chart; C, the partial control IA HScore QC chart.



3.3.1 Using IA HScore for positive controls

Final and JGH CLMA control blocks were examined as an alternative to the variability observed between and within tissue controls. There were 20 “Final” CLMA slides stained for ER, PR and HER2 and analyzed with IA. There were 54 JGH slides stained for HER2 and analyzed with IA. HScore QC charts were created for all slides for each IHC stain. Differences in SD from the mean HScore were compared for tissue and cell line cores of similar HScore with an F-test.

3.3.1.1 Practical application of cell-line based controls in QC charts

Final CLMA cell line QC charts were comparable in performance (similar trending patterns) to tissue QC charts. An observed difference in SD between cell line and tissue cores was statistically significant for some, but not all, of these comparisons. An example QC chart for a low ER HScore cell line core (MDA-MB-435), a medium PR HScore cell line core (MCF-7) and high HER2 HScore cell line (AU-565) are shown in Figure 10. This is the first direct observation of FFPE breast cancer cell line control QC charts, demonstrating their use as daily positive controls in QA.

Normal laboratory practice is to prepare many control slides at once and store unstained slides at -80°C. CLMA slides were cut at the clQc Program Laboratory. One was immediately stained for ER and a second slide was stored at -80°C for 30 days and then stained for ER. There were no significant measured decreases in ER HScore values for any cell line (paired *t*-test, Table 10). This preliminary work demonstrates that CLMA control slides can be sectioned and stored in the laboratory over time, in the same fashion as current controls, without sample degradation.

3.3.1.1.1 Reproducibility and homogeneity of cell lines

HScore values of duplicate cell line cores in “Final CLMA” were compared to each other by an unpaired *t*-test. No significant difference was observed between the HScores of duplicate cell line cores *except* for SKBR-3 for PR, and AU-565 and SKBR-3 for HER2. These differing cell lines had either extremely low mean HScore (PR SKBR-3, 14) or high mean HScore (HER2 AU-565, 263; HER2 SKBR-3, 269).

Cell line QC charts showed IA reproducibility for multiple cores of the same cell line

Figure 10: QC Charts for Daily IA HScore Results from Final CLMA Cell Line Cores. HScore values of twenty consecutive IHC-stained cell line cores were plotted to create a QC chart. Orange lines denote two SD from the mean HScore. Red lines denote 3 SD from the mean HScore. Values falling outside of 2 SD range indicate a need for recalibration of method and/or instrument. Values outside of 3 SD indicate the need for repeated testing because such results are unacceptable. A, low HScore core; B, medium HScore core; C, high HScore core.

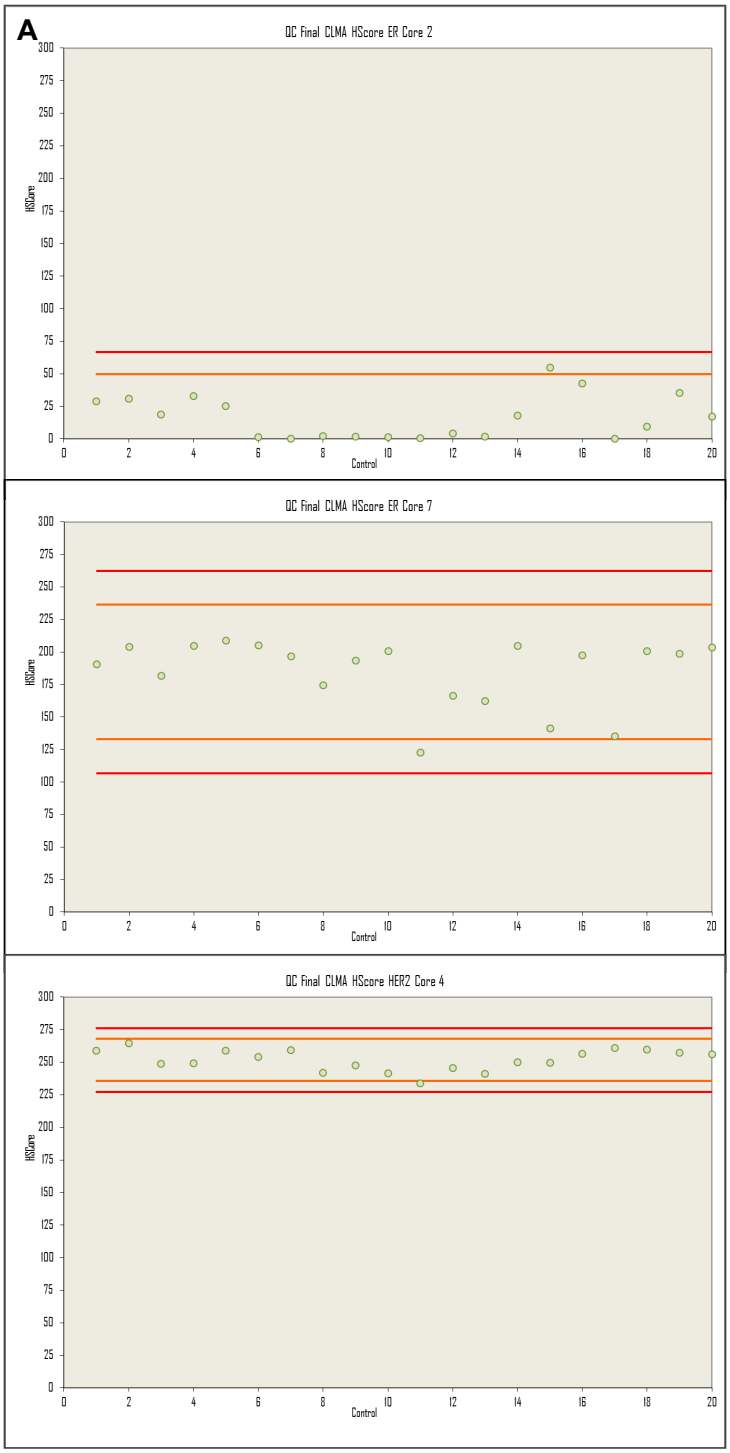


Table 10: Effect of -80°C Storage on IHC Staining of FFPE CLMA Slide Sections. HScore[a] of two slides, one stained by VGH immediately after preparation and one after 30 days storage at -80 °C.

Cell Line ID	0 days	30 days @ -80 °C
AU-565	63	71
BT-474	92	136
HBL-100	67	73
HS-578T	72	62
MCF-7	250	269
MDA-MB-231	45	87
MDA-MB-435	38	69
SKBR-3	(damaged)[b]	61
T47D	198	204

[a]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

[b]Damaged core was incorrectly evaluated by IA to have an erroneously high HScore (156).

over time when measured with IA. Trending and minimal variation in QC charts from two cores of the same intermediate cell line can be observed in Figure 11. This variation is reduced in lower staining cell lines, as shown in Figure 12. Cell lines with high HScores had the smallest SD. Comparison examples of duplicate high cell line core QC charts are in Figure 13. Intermediate HScore controls showed more visible variation over time, resulting in better visualization of shifts and trends in staining from day to day. In contrast, demonstration of biological homogeneity by this method cannot be observed for tissues, because no two samples are alike. For all cores, any HScore values falling outside of ±2 SD range indicate a need for recalibration of method and/or instrument. Values outside of 3 ± SD indicate the need for repeat testing because such results are unacceptable.

3.3.1.2 On-slide tissue and cell line controls in ER, PR and HER2 IHC

3.3.1.2.1 Final CLMA and JGH CLMA

Cell line QC charts had smaller SD than tissue QC charts. For this reason, cell lines controls were more discerning for the measurement of variation in system calibration. Consequently, a number of samples that were borderline-acceptable in cell line QC charts were acceptable by tissue QC charts. A comparison of cell lines and tissues HScores in Final TMA is found in Table 11. The CLMA demonstrated that SCH Laboratory is performing well, as no cores failed QC chart evaluation by being outside of ± 3 SD from the mean during this study (data not shown).

JGH selected two cell lines for visual verification of strongly positive (score, 3+) and low/negative (score, 0) HER2 staining. QC Charts from the cell line cores were matched with charts from two tissue cores of similar stain intensity (low/negative and strong). These results are shown in Figure 14. The low-staining cell line and tissue cores had HScore SDs of ± 0.5 and ± 0.2, respectively. The difference in low-staining core SD was statistically significant (*F*-test, $p<0.01$), with the HScore SD being slightly smaller for the cell line core. These small SD ranges were not sufficiently discriminatory for QC chart calibration use, even when graphed with a reduced *Y*-axis scale (data not shown). However, both negative- and strong- staining tissue and cell line samples produced very small QC chart SD ranges. Such small QC chart SD ranges are undesirable in

Figure 11: Reproducibility of Medium HScore Cell Line Core QC Charts From the Final CLMA. QC charts of PR HScore IA values from two cores of MCF-7 cell lines in Final CLMA control slides. Slides are from twenty consecutive IHC runs at SCH. Orange lines denote two SD from the mean HScore. Red lines denote three SD from the mean HScore. A, CLMA core 7; B, CLMA core 18.

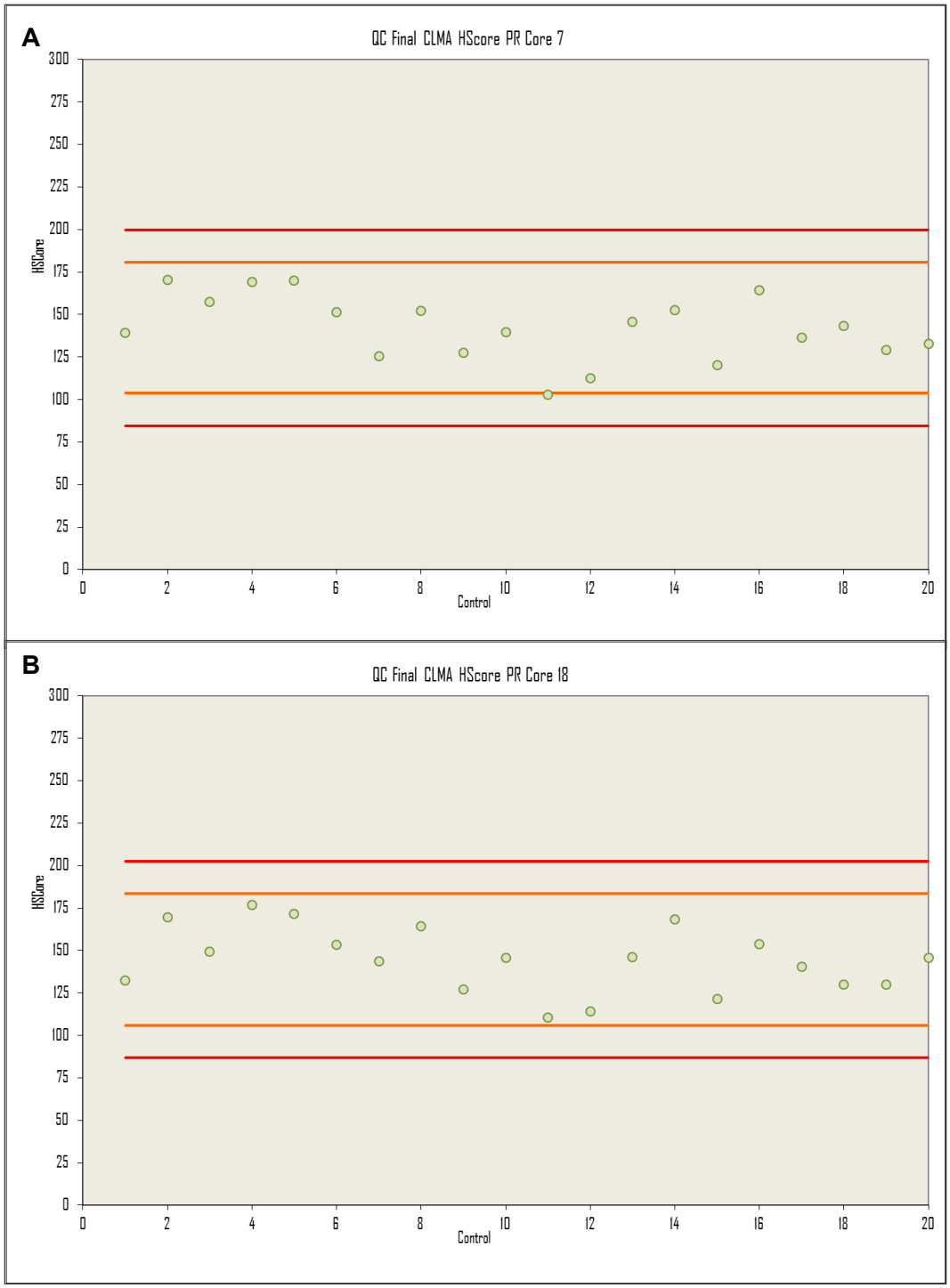


Figure 12: Reproducibility of Low HScore Cell Line Core QC Charts from the Final CLMA. QC charts of ER HScore IA values from two cores of MDA-MB-435 cell lines in Final CLMA control slides. Slides are from twenty consecutive IHC runs at SCH. Orange lines denote two SD from the mean HScore. Red lines denote three SD from the mean HScore. A, CLMA core 2; B, CLMA core 12.

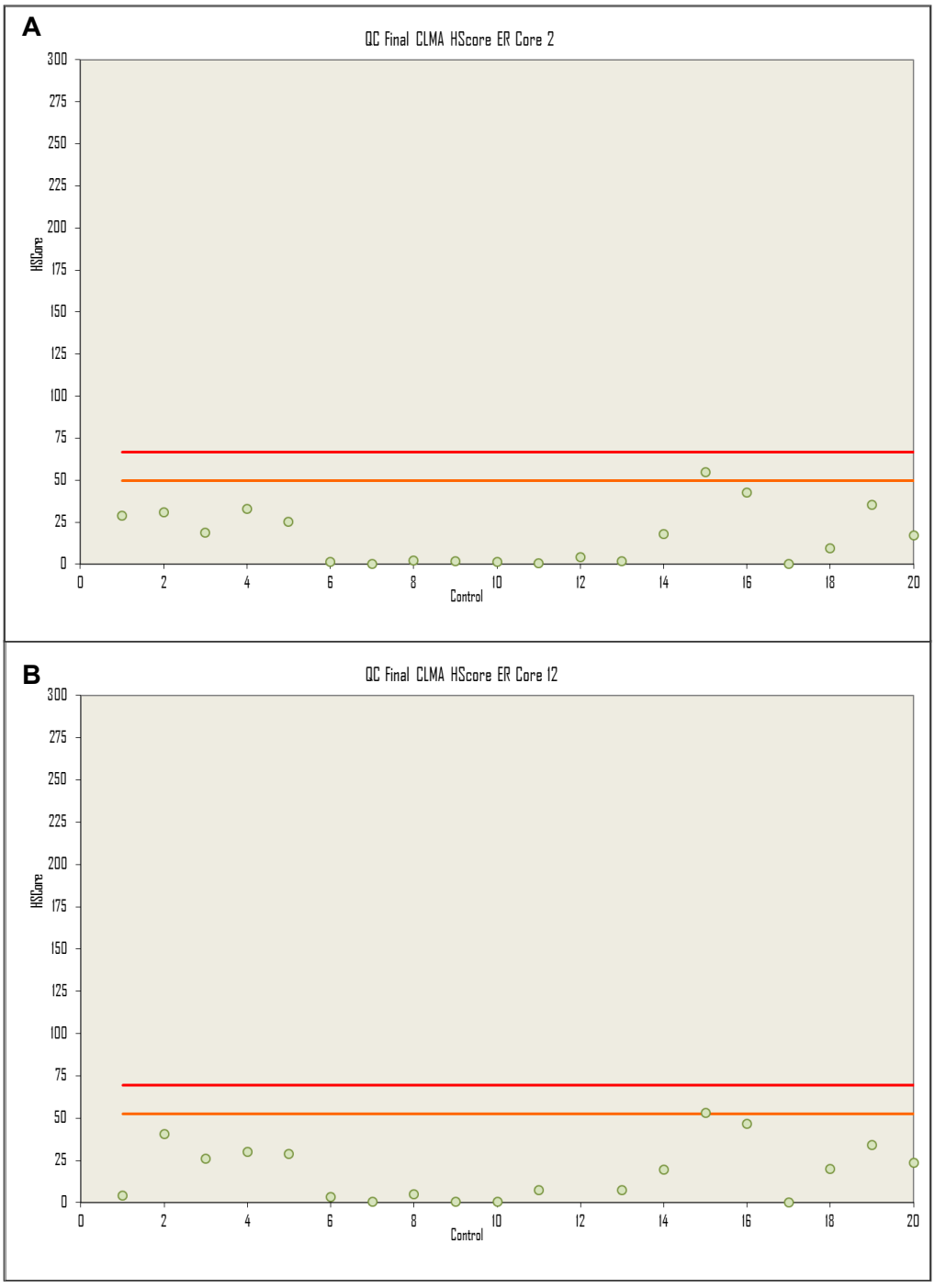


Figure 13: Reproducibility of High HScore Cell Line Core QC Charts from the Final CLMA. QC charts of HER2 HScore IA values from two cores of AU565 cell lines in Final CLMA control slides. Slides are from twenty consecutive IHC runs at SCH. Orange lines denote two SD from the mean HScore. Red lines denote three SD from the mean HScore. A, CLMA core 3; B, CLMA core 13.

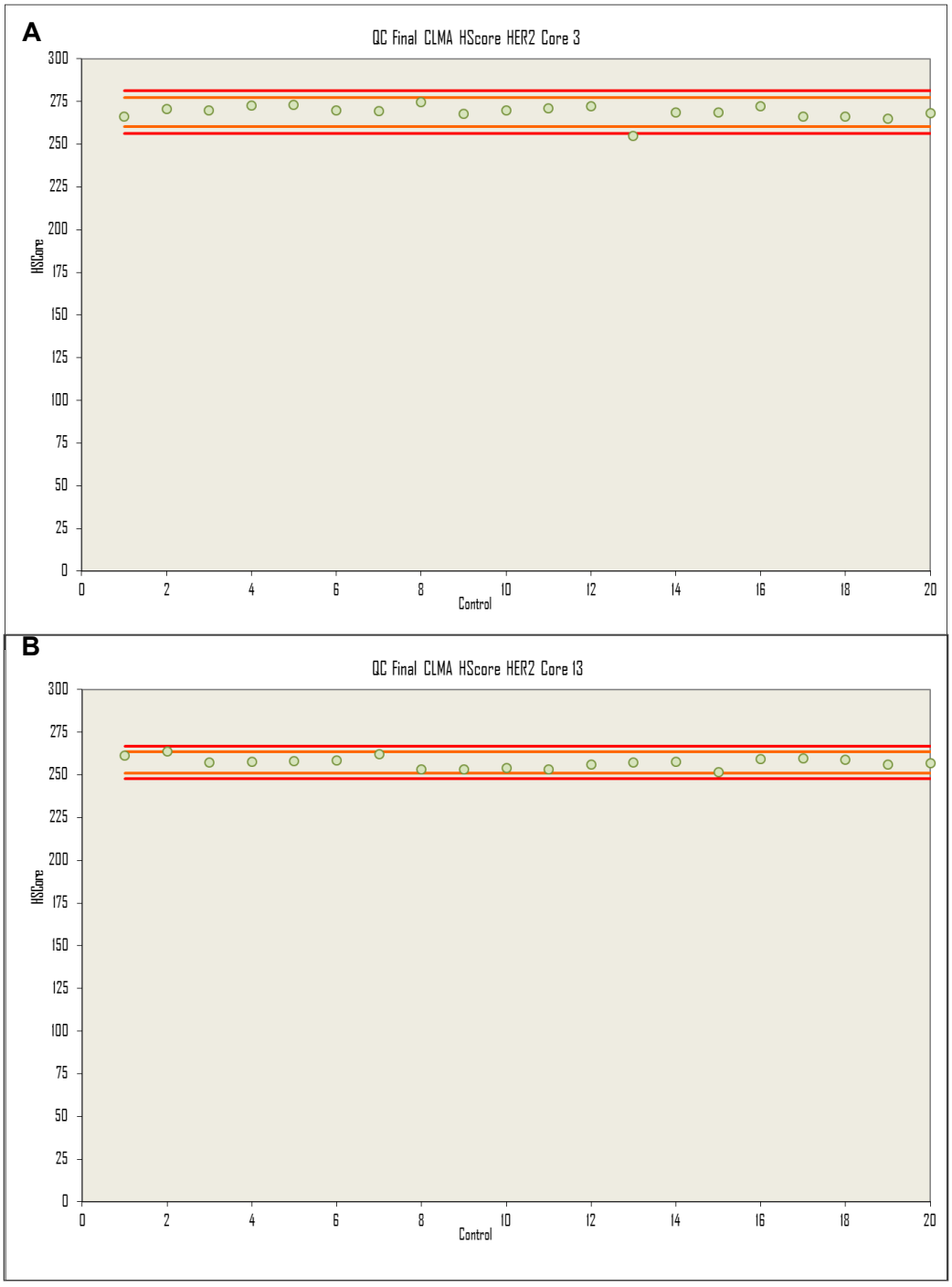


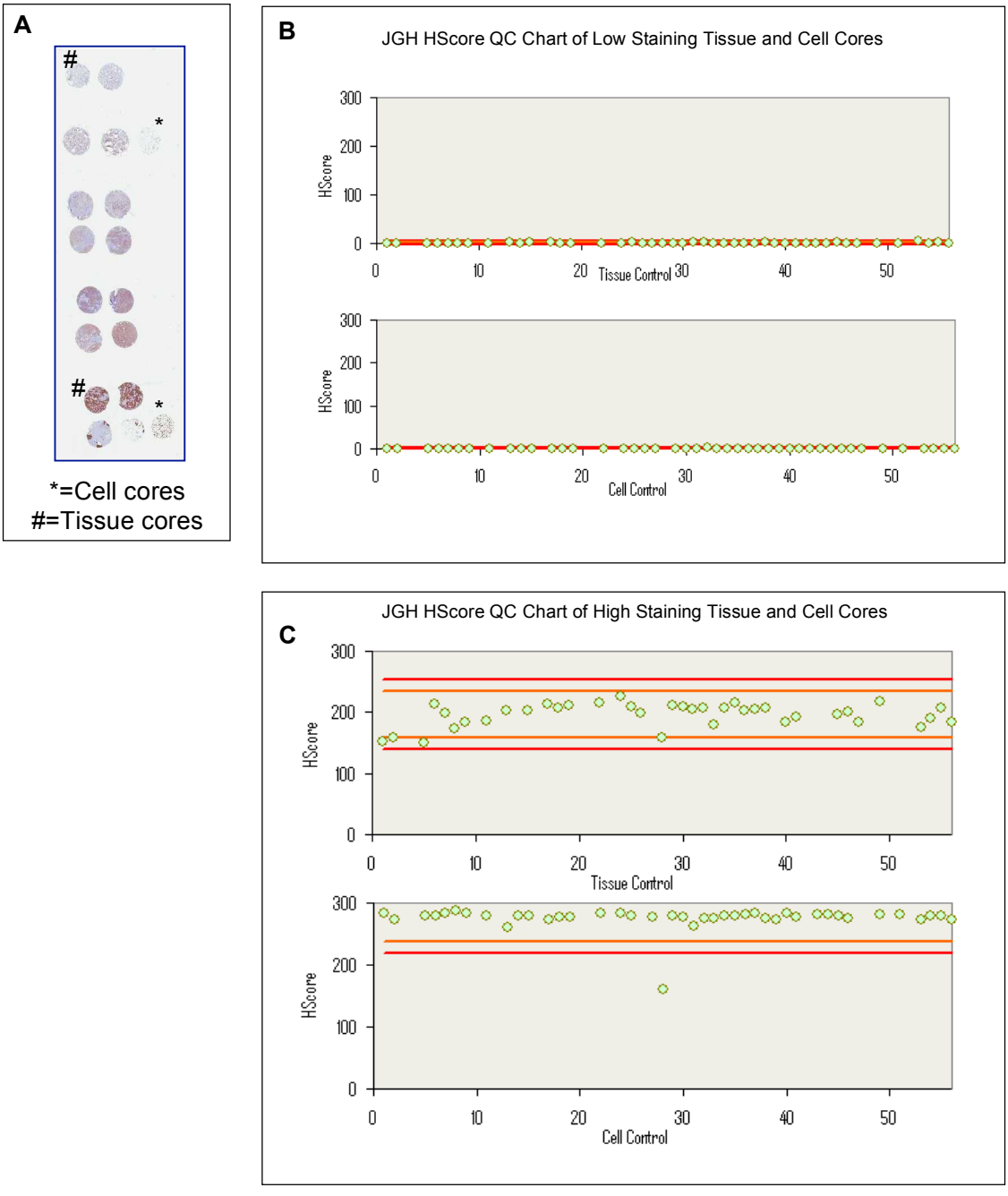
Table 11: Comparison of Cell Line and Tissue Core IA Mean HScore[a] [b] in the Final CLMA PT Control Block.

Cell Line	Core	ER	PR	HER2
AU-565	3	16 ± 4	17 ± 3	269 ± 1
	13	16 ± 3	16 ± 3	257 ± 2
BT-474	4	68 ± 10	241 ± 3	252 ± 2
	14	65 ± 9	233 ± 3	248 ± 2
HBL-100	5	8 ± 3	16 ± 6	1 ± 0.4
	15	13 ± 4	27 ± 7	2 ± 0.5
HS-576T	6	7 ± 2	15 ± 5	1 ± 0.2
	17	8 ± 2	17 ± 5	1 ± 0.2
MCF-7	7	185 ± 6	142 ± 4	12 ± 2
	18	188 ± 7	145 ± 4	7 ± 2
MDA-MB-231	1	16 ± 4	23 ± 5	2 ± 1
	11	22 ± 4	25 ± 5	1 ± 0.5
MDA-MB-435	2	16 ± 4	21 ± 5	0.1 ± 0.03
	12	17 ± 4	22 ± 4	0.1 ± 0.2
T47D	8	158 ±10	286 ± 1	131 ± 6
	19	163 ± 10	279 ± 1	153 ± 7
SKBR-3	9	14 ± 3	14 ± 3	268 ± 6
	20	16 ± 3	14 ± 3	270 ± 1

Tissue core	ER	PR	HER2
21	21 ± 3	51 ± 2	4 ± 1
22	25 ± 2	74 ± 2	10 ± 1
23	31 ± 2	30 ± 2	7 ± 1
25	118 ± 3	59 ± 2	9 ± 1
26	8 ± 1	140 ± 3	61 ± 3
27	103 ± 4	7 ± 1	1 ± 0.3
28	89 ± 2	47 ± 2	0.7 ± 0.1
29	77 ± 5	86 ± 2	1 ± 0.3
30	205 ± 4	10 ± 1	3 ± 0.5
31	2 ± 0.2	211 ± 5	1 ± 0.2
33	218 ± 11	5 ± 1	1 ± 0.4
34	2 ± 0.4	20 ± 1	72 ± 5
35	24 ± 2	4 ± 1	0.1 ±0.3
36	14 ± 2	64 ± 2	90 ± 2
37	103 ± 6	4 ± 1	2 ± 1
38	61 ± 6	4 ± 0.4	40 ± 5
39	32 ± 2	8 ± 1	39 ± 5
41	4 ± 1	20 ± 1	22 ± 3
42	8 ± 1	5 ± 1	1 ± 0.4
43	87 ± 5	3 ± 1	0 ± 0
44	63 ± 4	3 ± 1	27 ± 4
45	5 ± 1	5 ± 1	15 ± 1

[a]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).
[b]Results are an averaged HScore ± SD slides from twenty consecutive IHC slides stained by the SCH Laboratory.

Figure 14: Selected Tissue and Cell Core HScore QC Charts From the JGH CMLA HER2 Controls. Fifty six consecutive controls were stained at JGH. The two cell line cores of low and high IHC-stain intensity (HScore = 0.3 ± 0.5 and 275 ± 19) were compared to the two tissue cores with the most similar IHC-stain intensity (HScore = 0.7 ± 1.1 and 194 ± 19). Orange lines denote two SD from the mean HScore. Red lines denote three SD from the mean HScore. Values falling outside of 2 SD range indicate a need for recalibration of method and/or instrument. Values outside of 3 SD indicate the need for repeated testing because such results are unacceptable. A, the control containing sixteen breast tissue cores and two cell line cores (MDA-MB-231 and AU-565); B, comparison of low staining tissue and MDA-MB-231 cell line QC charts; C, comparison of high staining tissue and AU-565 cell line core QC Charts.



calibration materials. JGH control cell lines were not run in duplicate so no statistical comment on cell line homogeneity can be made.

3.3.2 Functionality of cells and tissues as controls

When used in QC control charts, HScores of both cell lines and tissue cores showed trending in staining performance. Samples out of ± 2 SD range could be observed using both control types. Cell line QC cores demonstrated biological homogeneity in IHC IA. Cell line QC charts had smaller SD and thus higher sensitivity in detecting true variation in calibration of the IHC system for ER and PR. IHC performance can be effectively monitored with a group of control cell lines with more than one staining intensity.

3.3.3 Selection of FFPE cell line controls for EQA

Cell lines AU-565 and HS-578T, AU-565 and MCF-7, and MCF-7 were identified as good midrange, positive controls for ER, PR and HER2, respectively. Cell lines MCF-7, T47D and AU-565 were selected as strongly positive controls for ER, PR and HER2, respectively. Cell line MDA-MB-231 is a low/ negative control for ER, PR and HER2.

3.3.3.1 RT-qPCR of FFPE cell lines: ER, PR and HER2 gene expression

A given set of FFPE cell lines blocks were analyzed by RT-qPCR and used to create a CLMA, which was stained for ER, PR and HER2 by five cIQc Program reference laboratories. Stained slides were visually assessed by cIQc Program assessors and by IA. RT-qPCR results for the nine cell lines were compared to IA HScore and pathologist assessment (Table 12). HScore levels of ER, PR and HER2 correspond well with cIQc visual assessment and RT-qPCR values. The results add an additional level of confirmation to the IHC and IA-measured levels of ER, PR and HER2 in the cell line samples. Results also indicate that the IHC-staining process, in itself, does not degrade ER, PR and HER2 in the nine FFPE cell lines prepared in this research. The use of the cell lines for daily positive controls as well as for EQA PT and evaluation of these markers is therefore possible, and can be performed by using IA.

Table 12: Comparison of IA HScore[a], RT-qPCR[b] Transcriptional Expression Results and Expert Pathologist Scoring Results[c] for Individually Prepared Blocks of Nine FFPE Cell Lines[d] That Were IHC-Stained for ER, PR and HER2.

	Average HScore ± SD			RT-qPCR transcription results			Expert Assessment Scores		
Cell Line	**ER**	**PR**	**HER2**	**ER**	**PR**	**HER2**	**ER**	**PR**	**HER2**
AU-565	51 ± 24	49 ± 24	246 ± 36	0.1	<0.01	221.3	N	N	3
BT-474	69 ± 27	190 ± 28	248 ± 6	1.6	6.3	156.5	P	P	3
HBL-100	33 ± 27	26 ± 23	3 ± 3	<0.1	<0.1	2.8	N	N	1
HS-578T	45 ± 26	36 ± 23	1 ± 2	<0.1	<0.1	1.3	N	N	0
MCF-7	245 ± 22	168 ± 24	11 ± 9	8.9	0.5	2.6	P	P	1
MDA-MB-231	40 ± 23	34 ± 24	1 ± 1	<0.1	<0.01	1.9	N	N	1
MDA-MB-435	31 ± 16	33 ± 19	1 ± 1	0.8	0.4	0.1	N	N	0
SKBR3	37 ± 26	65 ± 60	260 ± 9	0.1	<0.1	156.5	N	N	3
T47D	188 ± 39	268 ± 6	72 ± 24	5.3	6.3	6.5	P	P	2

[a]HScores range from 0-300; HScore = (3*(3%+))+(2*(2%+))+(1*(1%+)).

[b]RT-qPCR quantifies gene transcriptional expression from RNA extracted from the FFPE cell lines. Transcription levels are assumed to correlate with expression of ER, PR and HER2 within the samples. Sample results are expressed as ng protein equivalents, calculated by a relative quantification method that uses a calibrated external efficiency curve. The normalized gene copy numbers were adjusted using 5 housekeeper genes. Duplicate sample results were averaged.

[c]Expert assessment results are the average of EQA scoring results of slides analyzed by cIQc Program evaluating pathologists. Red cell line table squares are scored as positively stained (P), and white cell line squares are scored as negatively stained (N). For Her2, scores of 0 and 1 are negative and 2 and 3 are positive.

[d]IA HScore Results are an average of analyzed IHC slides stained by five cIQc reference laboratories.

4. 0 DISCUSSION OF RESULTS

4.1 IA in IHC

4.1.1 Using IA in PT

In this project, IA provided an accurate, reproducible measurement of IHC staining within and between EQA assessment runs. IA, when used in both in-laboratory daily quality control and in external, multiple-laboratory PT (EQA), produced satisfactory evaluative results. I conclude that IA can be used to monitor IHC results in PT.

An important feature of IHC evaluation by IA is the removal of human imprecision in measuring IHC results as well as human bias. This project has demonstrated that IA measurement is an unbiased, reproducible/consistent tool for evaluating laboratory PT performance within and between different PT runs. This precise and reproducible measurement is an important improvement in both the performance evaluation of participating laboratories and the measurement of laboratory method calibration. This is particularly important for all Class II IHC tests for which evidence-based calibration is needed.

Additionally, this study has also demonstrated the value of HScore and LSRSR for PT. HScore reflects an absolute amount of specific staining by IHC. Although the IHC system is not linear, it is considered to reflect the total amount of protein/epitope of interest. Despite this, an absolute HScore measurement has limited value in QA as there are no defined absolute reference measurements (gold standards) for any particular intensity of staining obtained for current samples used for PT. Also, there is no specific clinical use for absolute measurements, as there are none that define any disease or condition. The primary goal of PT is to compare calibration of the participating laboratories with reference laboratory results. Therefore, IA HScore data is more useful for assessing the calibration of the IHC procedure when converted into a LSRSR for PT evaluation, as a LSRSR reflects a measure of calibration relative to a reference laboratory [161]. This is the first study to employ HScore for the calculation of the LSRSR, and to emphasize the value of a ratio and comparison to a designated reference value, rather than trying to assess HScore on its own. It is important to

emphasize that reference values will vary from one series of testing to another and there is no gold standard to provide an absolute HScore value that can be used for comparison between laboratories.

4.1.2.1 Boxplots, histograms and Levey Jennings plots in PT

EQA programs need a means for results/data presentation from their PT runs. This study has demonstrated how EQA programs can use histograms to show the relative performance of a laboratory to the total group of participating laboratories. This project also demonstrated the usefulness of IA-derived boxplots and histograms, including their being practical in presenting relative performance in IHC PT. Herein, for the first time, HScores and LSRSR values were used in the plotting results in Levey Jennings plots of both cell line and tissue reference cores with equal success. This study has demonstrated that graphical illustration of results achieved with low-, medium- and high-expressing control cores is an excellent way of explaining EQA results to, and ensuring proper evaluation of, participating laboratories.

4.1.2.2 Identification of informative cell lines and informative human tissue samples for monitoring of IHC calibration in PT

This study evaluated which cell lines and tissue samples will be most informative for use in PT. To do this, I first had to determine which LSRSR ranges would be most informative to detect performance differences between participating laboratories. Initial IA analysis showed discriminatory LSRSR ranges for CLMA cores with HScore values in the range of 25 to 125. Subsequently, HScores of 50 to 150 were identified as practical and equally acceptable. Most importantly, both "very high" and "very low" staining cores were observed to have less discriminatory value. Therefore, samples with such HScores should not be selected for use in PT.

This study also considered which tissue samples or cell line samples would perform best in detecting system variation for daily QC. It is not surprising that, similar to PT, the best positive controls samples to be used for QC are the same samples that are used for PT; samples that have discriminatory LSRSR ranges between 50 and 150 were the best. This was confirmed by the observation that the cell line controls for which the HER2 LSRSR was less diagnostic were the cell lines that had LSRSR values higher

than optimal (HScore >150). Therefore, I conclude that tissue or cell line samples selected for PT (EQA) and in-laboratory daily QC should have these diagnostically useful, intermediate HScore values whenever possible. I have also confirmed that when IA and LSRSR are employed, fewer PT samples are required than with expert assessment of PT results (which requires >30 samples) to yield that same information about laboratory performance in PT. This was only preliminary suggested in our previous study [161]. Our current, and previous, studies indicate that LSRSR is an important EQA and QC tool. I, therefore, recommend the wide application of LSRSR in diagnostic IHC.

4.1.2.2.1 IA as an evaluation tool in cIQc Run 13

In Run 13 EQA, by using IA and LSRSR, I have identified more unacceptably/suboptimal results than with cIQc expert analysis (56%, 32% and 32% *versus* 34%, 25% and 5% for ER, PR and HER2, respectively). This initial work indicates that LSRSR appears to be a more sensitive discriminator by providing information about IHC protocol calibration, while the results of expert assessment provide different information (sensitivity, specificity, clinical validation). Both types of results are important for clinical IHC laboratories. Therefore, IA-derived LSRSR for both reference laboratories and participating laboratories in PT provides high-resolution evaluation of protocol calibration. IA is best employed in conjunction with biologically homogeneous cell line samples with informative HScores (ie., 50 – 150).

A number of designated cIQc reference laboratories also had one or more LSRSR values that were outside of an acceptable LSRSR range. This means that there are reference laboratories that provide results with more variation than previously recognized. It also suggests that these reference laboratories may still require further optimization in their methodology to achieve better reproducibility of their results.

Laboratories that participate in cIQc PT submit their protocols along with their stained slides. Since individual methods of the participating laboratories are available, consecutive PT followed by IA evaluations will facilitate improvement of the protocols and correlation of quality results with best methods, to further define best methods for IHC.

The cIQc Program uses a sufficient number of tissue samples (≥ 30) to enable the calculation of concordance, specificity, sensitivity, and κ to provide an assessment of participating laboratory performance characteristics. Due to the fact that the cIQc Program uses TMAs containing clinically representative human tissue samples, every PT run for ER, PR, and HER2 IHC is an external validation of the specific IHC protocols used in these tests. Despite this, the cIQc Program design and methods cannot be used to monitor specific IHC test calibration or to compare the calibration of the participating laboratories to the calibration of the reference laboratories. This is one of the most important roles of PT.

cIQc expertly scored EQA runs require 30 or more cores for analysis, and require assessor time and travel. Otherwise, no calculations of concordance and inter-laboratory reproducibility/variation are possible. This study showed, for the first time, that LSRSR is better than expert analysis for comparing IHC protocol calibration of participant laboratories to measurements obtained by the reference laboratory(-ies). This superiority is due to the reduced number of samples used and the fact that expert assessment is not required (reducing required time and travel costs). LSRSR reduces the number of samples required for PT while decreasing the turn-around time for issuing results of PT to participating laboratories.

4.1.2.2.2 IA as a calibration tool in PT: Run 13 *versus* Special Run

LSRSR in PT is best used as a calibration tool. This study has demonstrated that cell lines are more homogeneous in the expression of proteins than the human tissues currently used for either PT or QC. This is important if PT aims to monitor calibration of protocols. Human tissue samples are not very suitable for monitoring of IHC calibration because the differences detected between two or more tissue sections can be entirely due to biological tissue heterogeneity rather than protocol/system variation. Therefore, tissue samples are not particularly suitable as PT samples for monitoring of calibration.

This study showed a significantly lower variation between different sections from blocks prepared from cell lines, in comparison to sections prepared from tissue blocks. IA, with consistent cell line controls, allows evaluation and monitoring of the IHC protocol calibration that is not possible with heterogeneous tissue controls. Importantly,

the high sensitivity of IA paired with LSRSRs enabled the detection of intervention in participating laboratories after they were informed that their results were suboptimal. In almost all laboratories that had such results, there was a shift in sensitivity and protocol calibration in the next round of testing, which was only detected by IA and LSRSR, but not with expert assessment. This study also demonstrated that the participating laboratories who received excellent results did not show any significant evidence of change in calibration. Therefore, I also conclude that an LSRSR derived via IA can be used to detect whether there was an attempt to improve protocols and test performance after a laboratory receives unacceptable results in PT. Although statistically significant changes were not achieved for all three biomarkers, the ability to measure and demonstrate improvement and monitor protocol calibration of IHC is significant. This finely-tuned level of evaluation is simply not possible without the use of IA and LSRSRs. With successive cell line IHC PT runs, “optimal” HScore values and HScore ranges for each cell line will be better defined. This approach enables trend monitoring as well as correlation with protocol parameters to allow for the identification of suboptimal protocols and necessary method adjustment [20,161].

Study results identified several laboratories with suboptimal PT performance, via stained cores with suboptimal LSRSRs that were not identified as suboptimal by cIQc expert assessment. The question is whether this level of increased sensitivity of PT is useful for participating laboratories, and ultimately, for patient care. There is no intrinsic problem with increased sensitivity. The cut off levels for pass/fail could be changed or completely excluded from the reports, if necessary. It is advantageous for participating laboratories to have the opportunity to more precisely alter their protocol calibration towards an established gold standard. This gold standard is more informative and valuable than a simple pass/fail PT evaluation system. IA and LSRSR are superior to expert assessment because IA is objective and perfectly reproducible. There is also no cost for travel and accommodation of experts gathering for assessment from across Canada. This objectivity, reproducibility, and lower cost are all highly desirable in PT.

The role of expert assessment still remains very important for the quality of EQA programs. Human tumor TMA samples should continue to be used to provide information about clinical IHC sensitivity, specificity, and κ with reference results.

Human tissue TMA results are better evaluated by expert assessment because IA cannot reliably differentiate tumors from benign tissues. Experts should be able to identify and interpret this intrinsic, biological heterogeneity. An expert pathologist's role is also to select superior reference samples and gold standards (or reference laboratories) for calculation of LSRSR.

The level of calibration required for successful clinical practice could be identified through prospective studies that evaluated all of the samples from selected participating laboratories for a defined time period (e.g. 6 or 12 months), together with an appropriate reference laboratory. Only these semi-annual/annual audits would allow for final conclusions as to whether traditional expert assessment with human tissue TMAs or IA and LSRSR with cell line samples are more informative (or predictive of individual laboratory success with patient samples). It is generally understood that the PT and validation requirements traditionally applied in liquid-based assays (such as those frequently used in clinical chemistry laboratories) are not applicable for the evaluation of cell-based assays like IHC and flow cytometry with currently used methods. However, IA and LSRSR introduce a new methodology and a radical modification of current methods. I contend that using IA and LSRSR can move IHC towards the same QA standards as those used in liquid-based assays.

4.1.3 Use of IA in QC: HScore and QC charts in daily control evaluation

This project is the first demonstration of functional QC chart plotting of tissue and cell line IHC IA HScore data for in-laboratory/daily QA. Cell line-based IHC controls were more diagnostically useful, due to their superior functionality over tissue controls for ER and PR IHC. QA was optimal when QC charts were paired with biologically homogeneous, reproducible cell line control samples.

4.1.3.1 IA and IHC QA with current tissue controls

This study showed that current tissue-based positive IHC controls, when evaluated by IA and LSRSR, can also be used for creating QC charts. However, this study also quantifiably demonstrated the biological tissue heterogeneity of these controls. This heterogeneity is a potential source of variation that interferes with the primary purpose of controls. Controls need to be stable and consistently produce the same result. Both

tumor and benign human tissue show unacceptable variation in expression of an epitope/protein of interest. Although this variation is difficult to measure, document and follow with visual inspection, IA is excellent for this purpose. IA is sufficiently sensitive and reproducible, and is suitable for monitoring daily-run on-slide IHC controls for tissues.

The present study showed that, although tissue is biologically heterogeneous, this heterogeneity could be minimized by using larger tissue samples. Based on the results, the use of larger tissue samples is recommended, irrespective of the method for detection of results (visual inspection by microscope or IA). This is an effective approach in decreasing overall tissue heterogeneity.

This study also investigated different currently used tissues for their performance in daily QC. IHC-stain variation in control tissues from one day to another was examined using IA HScore. While it seems logical to use breast tumor tissue as a breast marker IHC control, this may not always be possible or practical due to variable tissue quality and availability. Examination of current control tissues from a number of hospitals did not conclusively determine a superior IHC control tissue for ER, PR and HER2. Breast tissue, endometrial tissue and a combination of endometrium and breast tissue all performed adequately. As discussed above, the size of the control was more important than tissue type as a parameter of sample quality. A group of the same tissue type with a range of known stain intensities is recommended by current national IHC guidelines [43,53].

This study also demonstrated that cell lines do not show biological heterogeneity for measured analytes from one level to another in a paraffin block. For this reason, cell lines are superior to human tissues as control samples. This has been concluded from the Final CLMA experiments where slides containing both FFPE tissue and cell line controls were stained in a single reference laboratory with previously demonstrated, optimal IHC PT performance. These successive runs showed that even in runs in which cell lines showed no variation from day to day, tissues occasionally showed unacceptable measurement. This tissue sample measurement variation may be falsely interpreted as system variation. This false suggestion of variation is unacceptable in a control sample. Therefore, these results suggest that cell line controls are a superior

alternative to human tissue controls. Cell line-based calibration models were confirmed as applicable daily positive controls for Class II IHC tests that could be monitored by IA with Levey Jennings QC charts.

4.1.3.2 FFPE QC standards

CAP, cIQc and NordiQC recommend that IHC laboratories run and monitor daily IHC controls, where percent positive tumor cells and stain intensity are recorded. Samples with various levels of expression of epitope of interest are recommended, while the use of a single sample with high expression of epitope of interest is discouraged, since it may be misleading [33,53,108]. Samples with very high expression levels do not provide any information about the level of system calibration [20,33,43].

It has been shown in many PT runs that a poorly performing laboratory is not able to identify their own poor performance most of the time because of the use of inappropriate controls in their daily QC [177]. Unfortunately, standardization of controls for clinical IHC has not been achieved yet and appropriate reference materials are still undefined. RT-qPCR results for the ER, PR and HER2 mRNA in the cell lines used in this study confirmed that IA appropriately measures IHC-stained levels of ER, PR and HER2 in the CLMA blocks, and that cell line manipulation and processing into paraffin blocks and subsequent manipulation by HIER does not significantly alter true biological levels of ER, PR and HER2 in FFPE tissue samples. This confirmation is important for CLMAs to be accepted as usable for IHC QA and PT in the future.

4.2 Characteristics of Cell Line Samples for ER, PR and HER2 IHC Controls

Study results suggest that by using IA, LSRSR, and selected cell lines as reference materials, standardization of IHC control samples can be done. The cell lines used need to have an HScore between 50 and 150 and need to be run as duplicate samples. Each paraffin block containing the cell line samples needs to be tested by a reference laboratory to produce reference results for each paraffin block. An increase in the density of cells in the cell pellets of the FFPE blocks to 4 X 10^7 cells in 300 µL is also desirable (i.e., compared to 2.0 X 10^7 cells in 300 µL as used in this thesis research). This higher cell density was suggested by expert cIQc pathologists as more desirable for use in IHC evaluation (personal communication, Blake Gilks and Emina Torlakovic).

While small variations in FFPE cell lines have been observed from batch to batch, these batch variations are not relevant when each set of results is correlated/calibrated using samples from within the same batch. In addition, each batch of cell lines will produce a large number of paraffin blocks with similar IHC IA results. This means that only a small number of different cell line batches are required to provide controls for a large amount of daily laboratory testing. It is entirely possible to produce a sufficient number of paraffin blocks for an end user for an entire year, from one cell line batch. For example, if a laboratory runs approximately 300 IHC tests for each of ER, PR, and HER2 per year, only two paraffin blocks would be required as controls for each test type.

I have identified useful reference/control cell lines for ER, PR and HER2 IHC. These cell lines are listed in Section 3.3 and Section 3.3.3. Previous studies looking at ER, PR and HER2[180] selected MCF-7, T47D and BT-474 as strongly staining cell lines for ER, PR and HER2, respectively. My results also observed this strong staining for these same cells. Dako currently uses MDA-MB-231 as a negative (Her2 score, 0) cell line and SK-BR3 as a strongly positive (Her2 score, 3+) cell line as a part of their HER2 IHC Herceptest™ [129]. My study results are in agreement with this research. However, I have also identified a number of intermediately staining breast cancer cell lines, which were shown to provide better resolution and allow more precise monitoring during IHC calibration. In addition, since both AU-565 and SKBR-3 cells similarly express HER2, only one of them needs to be used for making control samples. Also, both cell lines are difficult to culture, but the AU-565 cell line is faster growing and more robust, making it my recommended choice. Also, HBL-100 is a normal breast cell line[181] and was not selected as one of the recommended cell line control for any of the breast markers in this study. For this reason, HBL-100 cells also could be excluded to reduce labor and costs. From the cell lines tested, a CLMA control for ER, PR and HER2 IHC containing the cell lines AU-565, BT-474, HS-578T, MCF-7, MDA-MB-231, MDA-MB-435 and T47D is best.

4.2.1 Other considerations

IA is a very sensitive system. It has been previously demonstrated that the

thickness of a slide section must be carefully controlled because it can affect IHC-stain intensity. Thicker sections of FFPE cell lines have shown increased IHC staining [182]. For this reason, I recommend that the CLMA control slides used for calibration and QA be sectioned consistently, in the laboratories they are used in, to reduce this error. This is particularly important for the consistent IA evaluation used for PT, so that true variation in method or instrument performance can be measured.

The applications of this study are limited to FFPE samples and are restricted, at the moment, to ER, PR and HER2 IHC-stained human tissue. Study ideology can be applied on a larger scale to address other epitopes and proteins of interest if other cell lines are prepared in this manner and stained with other appropriate IHC methods and antibodies. QC charts could be used as suggested in clinical QC, either with large tissue controls or, ideally, with cell line controls in the optimal manner described. Further LSRSR analysis of future EQA runs will yield more data regarding the expanded applications of the results of this study. It is important to remember that the currently used clinical IHC methods and antibodies used in these methods are not fully standardized. As such, the limitations of the study are also related to the current quality of methodology used for IHC staining. As IHC improves, the results of this study will have wider applications.

4.3 Issues still remaining with IA

Since the start of this project, the practice of digital pathology has expanded to enhance training in, and the application of, digital pathology. Although this study shows that IA can be used for both PT and QC, a few remaining issues with IA need to be addressed.

IA cannot yet reliably identify tumor tissues or damaged samples. Although this is not relevant to cell line samples, it is entirely relevant for currently used human tissue controls. Pathologists use IHC to observe ER, PR, or HER2 protein expression in individual tumor cells and their subcellular location in positive cells. Expert assessment of human tumor samples prepared as TMAs for PT will exclude results from benign breast tissue. IA is unable to perform at that level in PT. Efforts to modify IA to reliably separate tumor and normal tissue were unsuccessful. Consequently, if IA is to be used

in PT with human tissue samples, the selection of samples for IA analysis must be done manually, prior to analysis. This is very labor intensive and is not possible to do, as it would make PT prohibitively expensive. I have partly addressed this problem by evaluating the effects the size of tested samples on detected error and true differences in protocol variation. IA scanning of human tissue samples was more consistent for larger samples. Smaller samples increased the error related to biological variations within the tissue (normal *versus* tumor). Smaller samples also increased the difference in measurement between different slide sections cut from the same paraffin block. Using cell lines instead of tissue as controls solved these two IA problems.

Although IA is reproducible and objective, section folding irrespective of source (cell lines or human tissue samples), will produce erroneous results. Therefore, any unusual IA LSRSR results must be examined as to their cause. I recommend post-scanning evaluation, because visual scoring of IHC samples is time consuming and it is impractical to screen every core on every slide prior to IA. This applies to both control samples in daily QC and samples used for PT. In this project, any unusual LSRSR results were easily clarified and corrected using post-scanning analysis. Currently, 40 tissue cores are used in PT for ER, PR and HER2 IHC, which could be replaced by a minimum of 3 cell line samples in the future. This would greatly simplify this post-scanning visual inspection for core folding because of the much shorter time needed.

4.3.1 Advances in IA and improvements in diagnostic IHC

The use of IA and digital microscopy in IHC will increase as the technology becomes more widely available and the cost of it decreases. However, until IHC methodology is standardized, it may be impractical to define and apply IA in a strictly quantitative manner. One of the most important messages of this work is that standardization of the overall, clinically applied, IHC is not possible without standardization of both the control samples used for both daily QC (with patient samples) and the samples used for PT (to compare each testing center with other testing centers and reference values). IA can help these developments.

4.4 Qualitative and Quantitative IHC: Are we there yet?

All IHC is currently qualitative, irrespective of how the results of IHC are used. Basically, the results are designated as positive or negative. There is a traditional, non-scientific cut off point of 10% cells with staining, used as a cut off point for interpreting IHC results as "positive". This is considered "qualitative IHC" and is used by pathologists for the interpretation of most Class I IHC tests that are employed to determine cell differentiation. For example, if the tumor is "positive" for cytokeratin, this favors the diagnosis of carcinoma *versus* the diagnosis of lymphoma. Most quantitative IHC is reduced, in clinical practice, to semi-quantitative scoring systems. The most frequent one is 0 to 3+. No cut off points are precisely defined and this system is not considered to be highly reproducible. Most recently, higher precision was recommended for Class II IHC tests with great consequence for patient treatment. In particular, ASCO/CAP guidelines for ER and PR IHC testing recommend a cut off at 1% weakly positive cells as a "positive result", declaring a patient eligible for hormonal therapy. The same recommendations ask for the use of HScore, or a similar scoring system, in which both intensity of staining and the percentage of positive cells is reported [43,45]. Additional methods have been developed for HER2 scoring since the start of this project [183].

One would assume that appropriate control systems (adequate for detecting errors at clinically significant cut off points) are available for daily QC of IHC testing. This is not the case. The recommended controls are generally a combination of tumor and benign tissues with no defined levels of expression. There is a discrepancy between the recommendation for precise evaluation of patient results and the lack of standardization of controls for QC. For these reasons, EQA programs that provide PT (like the cIQc Program) are seeking for answers to this problem.

This project has provided evidence of how to make better control samples together with recommendations for IA of the results. This will be very helpful, if applied in clinical practice. Also, my recommended evaluation with LSRSR enables the use of relatively inexpensive cell lines. The absolute levels of measured epitope in these useful cell lines is not relevant, as long as they are in the high resolution (intermediate) HScore range and a "gold standard" reference result for the each particular batch of cell lines is available. This applicable "gold standard" may be dictated by a provincial health

authority (for example, provincial reference laboratory results or national reference laboratory results) and was not a subject of this study.

4.5 Future prospects for IA, LSRSR and cell lines

The cIQc Program is eager to use CLMAs (as defined in this work) and LSRSR in their future ER, PR and HER2 IHC EQA evaluations, as well as other Class II PT. The cIQc Program has the facilities and expertise to culture large batches of cells for the preparation of FFPE cell line controls and provide “gold standard” results for these cultured cell lines. It is vital to have a single, consistent source of cell line control materials for clinical laboratory use. Although it is entirely acceptable for any center to develop these controls, many laboratories do not have the time, skills or funding required to reproducibly create FFPE CLMA controls or determine “gold standard” results for them. Both of these parameters are important and necessary, for reasons previously discussed. If a single provider of controls for PT and QC could be established, Canada would be the first country in the world to have standardized control samples for clinical/diagnostic IHC.

Optimally, CLMAs, IA, and the calculation of LSRSR would be combined as follows in daily QC for any laboratory that performs clinical IHC testing. A CLMA or cell line mixture control paraffin block (prepared commercially or by a designated center of excellence) would have the first cut slide stained by the designated reference laboratory and method. The remaining block would provide about 200 sections/control samples to be used in the diagnostic IHC laboratory for daily QC. The calculated LSRSR would be recorded on the Levey Jennings plots for each IHC test. This is very similar to currently run control samples in clinical chemistry laboratories. This approach enables proper calibration with a reference laboratory, as well as the daily monitoring of performance trends within a given diagnostic laboratory. The so-called "Westgard rules" [113] for monitoring controls could be evaluated in further studies for their applicability to this system.

A mixture of cells prepared as a cell block could also be considered for control sample preparation, which would obviate the need for creating a composite cell line microarray. A future study could be to create and evaluate a FFPE cell pellet control

containing a mixture of two cell lines. The preparation of this carefully quantified 50:50 mixture would require precise culturing of pre-selected cell lines by an experienced technologist. A scoring method for percentage of total cells staining positive denoting "successful" staining would need to be defined for scoring by IA or visual assessment. A negatively-staining, negative control cell line would be paired with an intermediately-staining, discriminatory cell line, with an optimal combination selected for each IHC epitope of interest. A strongly positive cell line control would not be necessary, because I have shown that such cell lines are not discriminatory for IHC calibration or evaluation. Drawing from my study results, I suggest the following combinations. The MDA-MB-231 cell line would function as a negative control for all three breast markers. Based on expert visual assessment in this project, I suggest the discriminatory cell line HS-578T or AU-565 for ER and PR, and cell line T47D for HER2, respectively. For IA assessment I recommend BT-474, MDA-MB-435 and MCF-7 for ER, PR and HER2 respectively, based on their HScore values in this project. For initial evaluation, paraffin blocks containing these mixtures could have cores added from individual FFPE cell lines (and/or known validated tissues) so that a comparison with current controls could be made. These controls could be evaluated as part of a cIQc breast cancer marker EQA run, to determine whether these mixtures would accomplish, essentially, the same PT control function as a CLMA, with reduced costs and preparation time.

I also envision a partnership with national and international IHC suppliers. Companies like Dako and Roche may be interested in selling these standardized controls on their own, or as a part of an entire IHC quality package, where instrumentation, reagents, controls and evaluation methods could be combined to optimize the entire procedure (in a manner similar to other clinical diagnostic tests). The optimized cell line samples, used as proposed here, would much cheaper than currently available cell line controls [129]. The price would be lower because this proposed system does not require an extremely precise monitoring of the absolute levels of the analyte of interest. The analyte only needs to be expressed in the cell lines in the approximate HScore linear range, which is wide (50 – 150). The innovative use of LSRSR is the key to this approach, because it is not critically important that clinical laboratories know what the absolute IHC levels are - merely how their results compare to designated “gold

standard” (thus making LSRSR very important). Additionally, this innovative use of IA enables the daily monitoring of controls on Levey Jennings plots, which is critical for detecting unacceptable system variation and errors. This is the first suggestion that PT (an important component for laboratory accreditation) can be standardized with QC charts. Connecting daily QC and clinical IHC performance to EQA evaluation makes this prospective method very clinically relevant.

4.6 Study Conclusions

1. IA can be used in diagnostic IHC PT, for EQA, to rank individual laboratory performance (in comparison to reference laboratory performance), using HScore and LSRSR. This method is more sensitive and more cost-efficient than currently used expert assessment methods.

2. A small number of cell line-based calibration samples are applicable in IHC PT and can be used in place of current TMA models, as they provide the same or more information than currently used TMA models for the calibration of IHC protocols. However, they cannot replace TMA samples for IHC validation. A combined cell line-based and human TMA sample-based methodology is recommended for use by EQA programs in order to provide both monitoring of calibration and a validation of IHC protocols.

3. Cell line-based calibration models are applicable as daily positive controls for Class II IHC tests. This type of controls can be evaluated by IA and monitored by using LSRSR and Levey Jennings-QC charts. IA QC charts allow precise observation trending in IHC performance, particularly when paired with cell line control samples.

Using IA with LSRSR to evaluate cell line controls is critical in the optimization and calibration of IHC methodology. Other cell lines can be evaluated as possible controls for other these and other Class II biomarkers. Through the diligent work of the cIQc, NordiQC and CAP Programs, IA has the potential to improve breast cancer diagnosis and treatment in Canada and around the globe.

5.0 REFERENCES

1. Cameron Commission of Inquiry. Cameron Commission of Inquiry on Hormone Receptor Testing. *Volumes 1-3*. 2009;1. Available at: www.cihrt.nl.ca/. Accessed May 5, 2012.

2. American Cancer Society. American Cancer Society Website. http://www.cancer.org. Accessed June 16, 2010.

3. Rosen PP. *Rosen's Breast Pathology*. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2009.

4. Ercan C, van Diest P, Vooijs M. Mammary development and breast cancer: the role of stem cells. *Current Molecular Medicine*. 2011;11(4):270-85.

5. Pelekanou V, Leclercq G. Recent insights into the effect of natural and environmental estrogens on mammary development and carcinogenesis. *The International Journal of Developmental Biology*. 2011;55(7-9):869-78.

6. Fridriksdottir AJR, Petersen OW, Rønnov-Jessen L. Mammary gland stem cells: current status and future challenges. *The International Journal of Developmental Biology*. 2011;55(7-9):719-29.

7. Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, Hayes DF, Kleer C, O'Malley PP, Page DL, Smith BL, Tan LK, Weaver DL, Winer EP. Members of the Cancer Committee, College of American Pathologists. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Archives of Pathology and Laboratory Medicine*. 2009;133:1515-38.

8. Francis GD, Dimech M, Giles L, Hopkins A. Frequancy and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by immunohistochemistry in Australia: results of the RCPA Quality Assurance Program. *Journal of Clinical Pathology*. 2007;60:1277-1283.

9. Cuadros M, Villegas R. Systematic review of HER2 breast cancer testing. *Applied Immunohistochemistry and Molecular Morphology*. 2009;17(1):1-7.

10. Terry J, Torlakovic EE, Garratt J, Miller D, Köbel M, Cooper J, Bahzad S, Pilavdzic D, O'Malley F, O'Brien AE, SenGupta S, Alport E, Têtu B, Knight B, Pettigrew NM, Berendt R, Wolber R, Trotter MJ, Riddell RH, Gaboury L, Elms F, Magliocco A, Barnes P, Gown AM, Gilks B. Implementation of a Canadian external quality assurance program for breast cancer biomarkers: an initiative of Canadian Quality Control in immunohistochemistry (cIQc) and Canadian Association of Pathologists (CAP) National Standards Committee/Immunohistoc. *Applied Immunohistochemistry & Molecular Morphology*. 2009;17(5):375-82.

11. Rhodes A, Jasani B, Barnes DM, Bobrow LG, Miller KD. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. *Journal of Clinical Pathology*. 2000;53(2):125-30.

12. Rhodes A, Jasani B, Balaton AJ, Barnes DM, Anderson E, Bobrow LG, Miller KD. Study of interlaboratory reliability and reproducability of estrogen and progesterone receptor assays in Europe. Documentation of poor reliability and identification of insufficient microwave antigen removal time as a major contributory element of unrelia. *American Journal of Cinical Pathology*. 2001;115:44-58.

13. Wells C a, Sloane JP, Coleman D, Munt C, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, Bussolati G, Connolly CE, Dervan P, Drijkoningen M, Ellis IO, Elston CW, Eusebi V, Faverly D, Heikkila P, Holland R, Jacquemier J, Lacerda M, Martinez-Penuela J, De Miguel C, Peterse JL, Rank F, Reiner A, Saksela E, Sigal-Zafrani B, Sylvan M, Borisch B, Cserni G, Decker T, Kerner H, Kulka J, Regitnig P, Sapino A, Tanous AM, Thorstenson S, Zozaya E. Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union--an inter-laboratory study. *Virchows Archiv : an international journal of pathology*. 2004;445(2):119-28.

14. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube SE, Somerfield MR, Hayes DF, Bast RC. American Society of Clinical Oncology 2007 Update of Recommendations for use of tumor markers in breast cancer. *Journal of Clinical Oncology*. 2007;25:5287-5312.

15. NordiQC. *Assessment Run B3 2007; estrogen receptor alpha (ER)*. 2007.

16. NordiQC. *Assessment Run B4 2007; HER-2*. 2007.

17. Lapidus RG, Nass SJ, Davidson NE. The loss of estrogen and progesterone receptor gene expression in human breast cancer. *Journal of Mammary Gland Biology and Neoplasia*. 1998;3(1):85-94.

18. Dodson A, Ibrahim M. UK-NEQAS Recommended Best Methods, In: Run 75, the Breast Hormonal Module: PR. *Immunocytochemistry*. 2007;6:74-78.

19. Hanna W, O'Malley PP, Barnes P, Berendt R, Gaboury L, Magliocco A, Pettigrew NM, Robertson S, SenGupta S, Tetu B, Thompson T. Updated Recommendations from the Canadian National Consensus Meeting on HER2/neu testing in breast cancer. *Current Oncology*. 2007;4:149-53.

20. Wolff AC, Hammond MEH, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez E a, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF. American Society of Clinical Oncology/College of American

Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Journal of Clinical Oncology*. 2007;25(1):118-45.

21. NordiQC. *Nordic QC Recommended Best Methods, In "Estrogen Receptor Aplha (ER), Run B5 2008."* 2008.

22. NordiQC. *Nordic QC Recommended Best Methods, In "HER-2, Run B7 2009."* 2009.

23. Ibrahim M. UK-NEQAS Recommended Best Methods in: "Run 76, The Breast Hormonal MReceptor Module: ER." *Immunocytochemistry*. 2008;(6):3.

24. UK-NEQAS. UK-NEQAS Recommended Best Methods, In: "Run 76, The Breast HER2 Model." *Immunohistochemistry*. 2008;6:139-144.

25. Walker R, Bartlett J, Dowsett M, Ellis I, Hanby A, Jasani B, Miller K, Pinder S. HER2 testing in the UK: further update to recommendations. *Journal of Clinical Pathology*. 2008;6:818-24.

26. Yaziji H, Taylor CR, Goldstein NS, Dabbs DJ, Hammond MEH, Hewlett B, Floyd AD, Barry TS, Martin AW, Badve S, Baehner F, Cartun RW, Eisen RN, Swanson PE, Hewitt SM, Vyberg M, Hicks DG. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. *Applied Immunohistochemistry and Molecular Morphology*. 2008;16(6):513-20.

27. Walker R. Quantification of immunohistochemistry--issues concerning methods, utility and semiquantitative assessment I. *Histopathology*. 2006;49(4):406-10.

28. Seidal T, Balaton AJ, Battifora H. Interpretation and quantification of immunostains. *The American Journal of Surgical Pathology*. 2001;25(9):1204-1207.

29. Goldstein NS, Watkins S. Immunohistochemistry. *Current Protocols in Molecular Biology*. 2008;January(Chapter 14):Unit 14.6.

30. Taylor CR. Quality assurance and standardization in immunohistochemistry. A proposal for the annual meeting of the Biological Stain Commission. *Biotechnic and Histochemistry*. 1992;67:110-17.

31. Taylor CR, Shi S-R, Chaiwun B, Young L, Imam SA, Cote RJ. Strategies for improving the immunohistochemical staining of various intranuclear prognostic markers in formalin-paraffin sections: androgen receptor, estrogen receptor, progesterone receptor, p53 protein, proliferating nuclear antigen and Ki-67 antigen r. *Human Pathology*. 1994;25:263-270.

32. Taylor CR, Levenson RM. Quantification of immunohistochemistry--issues concerning methods, utility and semiquantitative assessment II. *Histopathology*. 2006;49(4):411-24.

33. Torlakovic EE, Riddell RH, Hewlett B, Banerjee D, El-Zimaity H, Glynn G, Pilavdzic D, Dawe P, Magliocco A, Barnes P, Berendt R, Cook D, Gilks B, Williams G, Perez-Ordonez B, Wehrli B, Swanson OC, Nielsen S, Vyberg M, Butany J. Best practice recommendations for standardization of immunohistochemistry tests. *Canadian Journal of Pathology*. 2009;1:14-25.

34. Ramos-Vara JA. Principles and methods in immunohistochemistry. In: Gautier J, ed. *Drug Safety Evaluation: Methods and Protocols. Methods in Molecular Biology*. Volume 691. New York, NY: Springer Science + Business Media, LLC; 2011:83-96.

35. Taylor CR. From microscopy to whole slide digital images: a century and a half of image analysis. *Applied Immunohistochemistry and Molecular Morphology*. 2011;19:491-493.

36. Kumar GL, Rudbeck L eds. *Dako Educational Guide: Immunohistochemical Staining Methods*. 5th ed. Carpenteria, CA: Dako; 2009.

37. Iyengar JN. Quality Control in the Histopathology Laboratory: An overview with stress on the need for a structured national external quality assessment scheme. *Indian Journal of Pathology and Microbiology*. 2009;52(1):1-5.

38. Taylor C. The total test approach to standardization of immunohistochemistry. *Archives of Pathology and Laboratory Medicine*. 2000;124:945-51.

39. Goldstein NS, Ferkowicz M, Odish E, Mani A, Hastah F. Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *American Journal of Cinical Pathology*. 2003;120(1):86-92.

40. Goldstein NS, Hewitt SM, Taylor CR, Yaziji H, Hicks DG. Members of Ad-Hoc Committee on Immunohistochemistry Standardization. Recommendations for Improved Atandardization of Immuinohistochemistry. *Applied Immunohistochemistry and Molecular Morphology*. 2007;15(124-33).

41. Qiu J, Kulkarni S, Chandrasekhar R, Rees M, Hyde K, Wilding G, Tan D, Khoury T. Effect of delayed formalin fixation on estrogen or progesterone receptors in breast cancer: a study of three different clones. *American Journal of Clinical Pathology*. 2010;134(5):813-819.

42. Arber D. Effect of Prolonged Formalin Fixation on the Immunohistochemical Reactivity of Breast Markers. *Applied Immunohistochemistry and Molecular Morphology*. 2002;10(2):183-186.

43. Hammond MEH, Hayes DF, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG, Lester S, Love R, Mangu PB, McShane L, Miller K, Osborne CK, Pail S, Perlmutter J, Rhodes A, Sosano H, Schwartz JN, Sweep FCG, Taube S, Torlakovic EE, Valenstein PN, Vaile G, Visscher D, Wheeler T, Williams RB, Watliff JL, Wolff AC. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *Archives of Pathology & Laboratory Medicine*. 2010;134:907-922.

44. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG, Lester S, Love R, Mangu PB, McShane L, Miller K, Osborne CK, Paik S, Perlmutter J, Rhodes A, Sasano H, Schwartz JN, Sweep FCG, Taube S, Torlakovic EE, Valenstein P, Viale G, Visscher D, Wheeler T, Williams RB, Wittliff JL, Wolff AC. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of Clinical Oncology*. 2010;28(16):2784-95.

45. Canadian Association of Pathologists. Canadian Association of Pathologists Website-Resources. :Resources. Available at: http://www.cap-acp.org/resources.cfm. Accessed March 6, 2012.

46. O'Leary TJ. Standardization in immunohistochemistry. *Applied Immunohistochemistry and Molecular Morphology*. 2001;9:3-8.

47. Leong AS-Y, Leong TY-M. Standardization in immunohistology. *Methods in Molecular Biology*. 2011;724:37-68.

48. Khoury T, Sait S, Hwang H, Chandresekhar R, Wilding G, Tan D, Kulkarni S. Delay to formalin fixation effect on breast biomarkers. *Modern Pathology*. 2009;22(11):1457-1467.

49. NordiQC. Nordic Immunohistochemistry Quality Control website. Available at: www.nordicqc.org. Accessed January 11, 2011.

50. Larsson L. Tissue preparation methods for light microscopic immunohistochemistry. *Applied Immunohistochemistry*. 1993;1:2-16.

51. Kothmaier H, Rohrer D, Stacher E, Quehenberger F, Becker KF, Popper HH. Comparison of formalin-free tissue fixatives: a proteomic study testing their application for routine pathology and research. *Archives of Pathology and Laboratory Medicine*. 2011;135(6):744-52.

52. Engel K, Moore H. Effects of Preanalytical Variables on the Detection of Proteins by Immunohistochemistry in Formalin-FIxed, Paraffin-Embedded Tissue. *Archives of Pathology and Laboratory Medicine*. 2011;135:537-543.

53. Hammond MEH, Hayes DF, Wolff AC. Clinical Notice for American Society of Clinical Oncology-College of American Pathologists guideline recommendations on ER/PgR and HER2 testing in breast cancer. *Journal of Clinical Oncology*. 2011;29(15):e458.

54. Leong AS-Y, Gilham PN. The effects of progressive formaldehyde fixation on the preservation of tissue antigens. *Pathology*. 1989;21:81-89.

55. Leong AS-Y. Fixation and fixatives. In: Woods A, Ellid R, eds. *Laboratory Histopathology - A complete Reference*. London England: Churchill Livingstone; 1994.

56. Grizzle WE, Grody WW, Noll WW, Sobel ME, Stass SA, Trainer T, Travers H, Weedn V, Woodruff K. Recommended policies for use of human tissue in research, education and quality control. Ad Hoc Committee on Stored Tissue, College of American Pathologists. *Archives of Pathology & Laboratory Medicine*. 1999;123(4):296-300.

57. Bell WC, Young ES, Billings PE, Grizzle WE. The efficient operation of the surgical pathology gross room. *Biotechnic and Histochemistry*. 2008;83(2):71-82.

58. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. *Modern Pathology*. 2008;21 Suppl 2:S8-S15.

59. Vyberg M, Torlakovic EE, Seidal T, Risberg B, Helin H, Nielsen S. Nordic Immunohistochemistry Control. *Croatian Medical Journal*. 2005;46:368-371.

60. Oyama T, Ishikawa Y, Hayashi M, Arihiro K, Horiguchi J. The Effects of Fixation , Processing and Evaluation Criteria on Immunohistochemical Detection of Hormone Receptors in Breast Cancer. *Breast Cancer*. 2007;14(2):182-188.

61. Sainte-Marie G. A paraffin embedding technique for studies employing fluorescence. *Journal of Histochemistry and Cytochemistry*. 1962;10:250-256.

62. Bosman F, Linderman J, Kuiper G, van der Wal A, Kreunig J. The influence of fixation on immunoperoxidase staining of plasma cells of intestinal biopsy specimens. *Histochemistry*. 1977;53:57-62.

63. Buorges M, Esterre P, Gurrett S, Grimaud JA. A simplified method of tissue processing for immunostaining with good preservation of antigen and morphology. *Cell Molecular Biology*. 1989;35:231-237.

64. Sato Y, Mukai K, Furuya S, Kameya T, Hirohashi S. The AMeX method: a multipurpose tissue processing and paraffin embedding method. *American Journal of Cinical Pathology*. 1992;140:775-779.

65. Bartheau P, Cazals-Hatem D, Meignin V, de Roquancourt A, Verola O, Lesourd A, Sene C, Brocheriou C, Janin A. Variability of immunohistochemical reactivity on stored paraffin slides. *Journal of Clinical Pathology*. 1998;51:370-374.

66. Jacobs TW, Prioleau JE, Stillman IE, Schnitt SJ. Loss of tumour marker immunostaining intensity on stored paraffin slides of breast cancer. *Journal of the National Cancer Institute*. 1996;88:1054-1059.

67. Steu S, Baucamp M, von Dach G, Bawohl M, Dettwiler S, Storz M, Moch H, Schraml P. A procedure for tissue freezing and processing applicable to both intra-operative frozen section diagnosis and tissue banking in surgical pathology. *Virchows Archiv : an international journal of pathology*. 2008;452(3):305-12.

68. Shi S-R, Cote RJ, Taylor CR. Antigen Retrieval Immunohistochemistry: Past, present and future. *Journal of Histochemistry and Cytochemistry*. 1997;45:327-343.

69. Bratthauer GL. Preparation of Frozen Sections for Analysis. In: Oliver C, Jamur M, eds. *Immunocytochemical Methods and Protocols*. 3rd ed. New York, NY: Humana Press; 2010:67-74.

70. Macintyre N. Unmasking genes for immunohistochemistry. *British Journal of Biomedical Science*. 2001;58:190-196.

71. Leong AS-Y, Leong TY-M. How does antigen retrieval work? *Advances in Anatomic Pathology*. 2007;14(2):129-131.

72. Anagnostou VK, Welsh AW, Giltnane JM, Siddiqui S, Liceaga C, Gustavson M, Syrigos KN, Reiter JL, Rimm DL. Analytic variability in immunohistochemistry biomarker studies. *Cancer Epidemiology Biomarkers and Prevention*. 2010;19(4):982-91.

73. Curran RC, Gregory J. The unmasking of antigens in paraffin sections of tissue by trypsin. *Experentia*. 1977;33:1400-1401.

74. Shi S-R, Tian Q. Development of an antigen retrieval technique for immunohistochemistry on archival celloidin-embedded sections. *Journal of Histochemistry and Cytochemistry*. 1993;41:1121.

75. Shi S-R, Liu C, Taylor CR. Standardization of immunohistochemistry for formalin-fixed, paraffin-embedded tissue sections based on the antigen-retrieval technique: from experiments to hypothesis. *Journal of Histochemistry and Cytochemistry*. 2007;55(2):105-9.

76. Shi S-R, Cote RJ, Young. Development of optimal protocols for antigen retrieval immunohistochemistry based on the effects of variation in temperature and pH. In: Bailey GW, Ellisman MH, Hennigar R, Zaluzec NJ, eds. *JMSA proceedings in microscopy and microanalysis*. New York, NY; 1995.

77. Kanai K, Nunoya K, Shibuya T, Nakamura M, Tajima M. Variations in effectiveness of antigen retrieval pretreatments for diagnostic immunohistochemistry. *Research in Veterenary Science*. 1998;64:57-61.

78. Dako. Dako Website. Available at: www.Dako.com. Accessed March 4, 2011.

79. Boenisch T. Can a more selective application of antigen retrieval facilitate standardization in immunohistochemistry? *Applied Immunohistochemistry and Molecular Morphology*. 2004;12(2):172-6.

80. Myers J. Antigen Retrieval: a review of commonly used methods and devices. *Medical Laboratory Observer*. 2006;38:10-15.

81. McCarty Jr. KS, MIller Jr. LS. Estrogen receptor analyses: correlation of biochemical and immunohistochemical methods using monoclonal and anti-receptor antibodies. *Archives of Pathology and Laboratory Medicine*. 1985;109:716-721.

82. Reiner L. On the chemical alteration of purified antibody-proteins. *Science*. 1930;72:483-484.

83. Eagle H, Smith D, Vickers P. The effect of combination with diazo compounds on the immunological reactivity of antibodies. *Journal of Experimental Medicine*. 1936;63(5):617-643.

84. Nakane PK, Pierce GB. Enzyme labelled antibodies: preparation and application for the localization of antigens. *Journal of Histochemistry and Cytochemistry*. 1967;14:929-931.

85. Sternberger LA, Hardy PH, Cuculia JJ, Meyer HG. The labeled antibody enzyme method of immunohistochemistry. Preparation and properties of soluble antigen-antibody complex (horseradish peroxidase-antihorseradish peroxidase) and its use in identification of spirochetes. *Journal of Histochemistry and Cytochemistry*. 1970;18:315-333.

86. Dorset BH, Joachim HL. A method for the use of immunofluorescence on paraffin embedded tissues. *American Journal of Cinical Pathology*. 1978;69:66-72.

87. Leong TY-M, Cooper K, Leong AS-Y. Immunohistology- past, present and future. *Advances in Anatomic Pathology*. 2010;17(6):404-418.

88. Torlakovic EE, Riddell R, Banerjee D, El-Zimaity H, Pilavdzic D, Dawe P, Magliocco A, Barnes P, Berendt R, Cook D, Gilks B, Williams G, Perez-Ordonez B, Wehrli B, Swanson PE, Otis CN, Nielsen S, Vyberg M, Butany J. Canadian Association of Pathologists-Association canadienne des pathologistes National Standards Committee/Immunohistochemistry: best practice recommendations for standardization

of immunohistochemistry tests. *American Journal of Cinical Pathology*. 2010;133(3):354-65.

89. Oliver C, Jamur MC. Overview of antibodies for immunohistochemistry. In: Oliver C, Jamur MC, eds. *Immunocytochemical Methods and Protocols*. 3rd ed. New York, NY: Humana Press; 2010:3-10.

90. Sinkovics J, Dreesman G. Monoclonal antibodies of hybridomas. *Reviews of Infectious Diseases*. 1983;5:9-34.

91. McCormick D, Yu C, Hobbs C, Hall PA. The relevance of antibody concentration to the immunohistological quantification of cell proliferation-associated antigens. *Histopathology*. 1993;22:543-547.

92. Bruns J, Hambridge M, Taylor CR. Intracellular immunoglobulins: a comparative study on three standard tissue processing methods using horseradich peroxidase and fluorochrome conjugates. *Journal of Clinical Pathology*. 1974;27:548-557.

93. Montero C. The antigen-antibody reaction in immunohistochemistry. *Journal of Histochemistry and Cytochemistry*. 2003;51(1):1-4.

94. Krenacs I, Krenacs L, Raffield M. Multiple Antigen Immunostaining Procedures. In: Oliver C, Jamur MC, eds. *Immunocytochemical Methods and Protocols*. 3rd ed. New York, NY: Humana Press; 2010:67-74.

95. Ventana. Ventana Medical Systems Website. Available at: www.ventanamed.com. Accessed March 5, 2011.

96. IHC World. IHC World Website. Available at: www.ihcworld.com/introduction.htm. Accessed March 4, 2011.

97. Lequin RM. Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical Chemistry*. 2005;51(12):2415-8.

98. Rimm DL. What brown cannot do for you. *Nature biotechnology*. 2006;24(8):914-6.

99. Leong AS-Y. Pitfalls in diagnostic immunohistochemstry. *Advances in Anatomic Pathology*. 2004;11(2):86-93.

100. Leong AS-Y. Quantitation in immunohistochemistry. Fact or Fiction? A discussion of variables that influence results. *Applied Immunohistochemistry and Molecular Morphology*. 2004;12:1-7.

101. Bogen S a, Vani K, McGraw B, Federico V, Habib I, Zeheb R, Luther E, Tristram C, Sompuram SR. Experimental validation of peptide immunohistochemistry controls. *Applied Immunohistochemistry and Molecular Morphology*. 2009;17(3):239-46.

102. Canadian Immunohistochemistry Quality Control. Canadian Immunohistochemistry Quality Council Website. 2009. Available at: www.ciqc.ca. Accessed March 9, 2011.

103. Linnet K, Boyd JC. Selection and Analytical Evaluation of Methods with Statistical Techniques. In: Burtis CA, Ashwood ER, Bruns DE, eds. *Teitz Textbook of Clinical Chemistry and Molecular Diagnostics*. 4th ed. St. Louis, Missouri: Elsevier Saunders; 2006.

104. Fetsch PA, Abati A. The Clinical Immunohistochemical Laboratory: Regulations and Troubleshooting Guidelines. In: Oliver C, Jamur MC, eds. *Immunocytochemical Methods and Protocols*. 3rd ed. New York, NY: Humana Press; 2010:67-74.

105. Thompson M, Ellison SLR, Wood R. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). *Pure and Applied Chemistry*. 2006;78(1):145-196.

106. Kalra J. *Medical Errors and Patient Safety: Strategies to reduce and disclose medical errors and improve patient safety*. Germany: DeGruyter; 2011.

107. Montagu D. *Accreditation and other External Quality Assessment Systems for Healthcare: Rreview of experience and lessons learned*. London; 2003.

108. Vyberg M. Prof. Vyberg: “Quality Assurance in IHC: The NordiQC Experience.” 2011. Available at: http://www.youtube.com/watch?v=IoZP1F5SxqM. Accessed April 26, 2012.

109. Shewhart WS. *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York; 1931:94, 147, 249–261, 290–291, 309.

110. Henry RJ. Use of the Control Chart in Clinical Chemistry. *Clinical Chemistry*. 1959;5(4):309-319.

111. Shewhart WS. *Statistical Method From the Viewpoint of Quality Control*. Washington, DC; 1939.

112. Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Journal of Clinical Pathology*. 1950;20:1059-1066.

113. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. *Clinical chemistry*. 1977;23(10):1857-67.

114. Carroll TA, Minnick HA, Carroll WE. Brief Communications: Probability and the Westgard Rules. *Annals of Clinical and Laboratory Science*. 2003;33:113-114.

115. Westgard JO, Groth T, Aronsson T, de Verdier CH. Combined Shewhart-cusum control chart for improved quality control in clinical chemistry. *Clinical chemistry*. 1977;23(10):1881-7.

116. Bishop J, Nix a B. Comparison of quality-control rules used in clinical chemistry laboratories. *Clinical Chemistry*. 1993;39(8):1638-49.

117. Westgard JO. Managing quality vs. measuring uncertainty in the medical laboratory. *Clinical Chemistry and Laboratory Medicine*. 2010;48(1):31-40.

118. Wadsworth HM. Statistical Process Control. In: Juran J, Godfrey AB, eds. *Juran's Quality Handbook*. 5th ed. New York, NY: McGraw-Hill; 1999:45.1-45.29.

119. Carey R. Constructing Powerful Control Charts. *Journal of the Association of Computing Machinery*. 2002;October:65-70.

120. Bogen SA, Sompuram SR. Peptides as Immunohistochemistry Controls. In: Shi S-R, Taylor CR, eds. *Antigen Retrieval Immunohistochemistry Based Research and Diagnostics*. Hoboken, NJ: John Wiley & Sons; 2010.

121. Sompuram SR, Kodela V, Zhang K, Ramanathan H, Radcliffe G, Falb P, Bogen SA. A novel quality control slide for quantitative immunohistochemistry testing. *Journal of Histochemistry and Cytochemistry*. 2002;50:1425-1433.

122. Gown AM. Tweaking and Nudging Towards Improved IHC Quality. *Applied Immunohistochemistry and Molecular Morphology*. 2009;17:363-365.

123. Aitio A, Apostolib P. Quality assurance in biomarker measurement. *Toxicology Letters*. 1995;77:195-204.

124. Hsu FD, Nielsen TO, Alkushi A, Dupuis B, Huntsman D, Liu CL, van de Rijn M, Gilks CB. Tissue Microarrays Are an Effective Quality Assurance Tool for Diagnostic Immunohistochemistry. *Modern Pathology*. 2002;15(12):1374-1380.

125. Regitnig P, Reiner A, Dinges H-P, Höfler G, Müller-Holzner E, Lax SF, Obrist P, Rudas M, Quehenberger F. Quality assurance for detection of estrogen and progesterone receptors by immunohistochemistry in Austrian pathology laboratories. *Virchows Archiv*. 2002;441(4):328-34.

126. Wasielewski R, Hasselmann S, Ruschoff J, Fisseler-Eckhoff A, Kreipe H. Proficiency testing of immunohistochemical biomarker assays in breast cancer. *Virchows Archiv*. 2009;453:537-543.

127. Michaud GY. The role of standards in the development and implementation of clinical laboratory tests: a domestic and global perspective. *Cancer Biomarkers*. 2005;1(4-5):209-16.

128. Barker PE. Material standards and reference materials for disease biomarkers : Needs assessment and development. *Cancer Biomarkers*. 2006;1(2005):207-208.

129. Dako. Guidelines for Scoring HercepTest TM - Gastric Guideline for Scoring of Control Cell Lines. 2008;0:4. Available at: http://www.dako.com/herceptest_scoring_guidelines_-_gastric_ihc_38647.pdf. Accessed June 5, 2012.

130. Vani K, Sompuram SR, Fitzgibbons PL, Bogen SA. National HER2 Proficiency Test Results Using Characterization of Laboratory Failures. *Archives of Pathology and Laboratory Medicine*. 2008;132:211-216.

131. Westgard JO. Personal View Internal quality control : planning and implementation strategies. *Annals of Clinical Biochemistry*. 2003;40:593-612.

132. Hollensead SC, Lockwood WB, Elin RJ. Errors in pathology and laboratory medicine: consequences and prevention. *Journal of Surgical Oncology*. 2004;88(3):161-81.

133. Juran JM, Blanton GA eds. *Juran's Quality Handbook*. 5th ed. New York, NY: McGraw-Hill; 1998.

134. QMPLS. Quality Management Program -Laboratory Services (Ontario laboratory assessment program). Available at: http://www.qmpls.org/eqa/eqa_path.html. Accessed February 9, 2011.

135. Manual ACS. *AJCC Cancer Staging Manual*. 7th ed. (Edge S, Byrd D, Carducci M, Compton C, eds.). New York, NY: Springer; 2009.

136. Bacus S, Flowers JL, Press MF, Bacus JW, McCarty, K S S. The evaluation of estrogen receptor in primary breast carcinoma by computer assisted image analysis. *American Journal of Clinical Pathology*. 1988;90:233-239.

137. Bloom K, Harrington D. Enhanced Accuracy and Reliability of HER-2/neu Immunohistochemical Scoring Using Digital Microscopy. *American Journal of Clinical Pathology*. 2004;121(5):620-630.

138. Bernardo V, Laurenco S, Cruz R, Monteiro-Leal L, Silvo L, Camisasca D, Farina M, Lins U. Reproducability of Immunostaining quantification and Description of a New Digital Processing Procedure for Quantitative Evaluation of Immunohistochemistry in Pathology. *Microscopy and Analysis*. 2009;15:353-365.

139. Turbin DA, Leung S, Cheang MCU, Kennecke H a, Montgomery K, McKinney S, Treaba DO, Boyd N, Goldstein LC, Badve S, Gown AM, van de Rijn M, Nielsen TO, Gilks B, Huntsman DG. Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue

microarray study of 3,484 cases. *Breast Cancer Research and Treatment*. 2008;110(3):417-26.

140. Slodkowska J, Filas V, Buszkiewski E, Trzeciak P, Wojciechowski M, Koktysz R, Staniszewski W, Breborowicz J, Rojo M. Study on breast carcinoma HER2/neu and hormonal receptors status assessed by automated images analysis systems: ACIS III (Dako) and ScanScope (Aperio). *Folia Histochemica et Cytobiologica*. 2010;48:19-25.

141. Kresak L, Micsik T, Kiszler G, Krenacs T, Szabo D, Jonas V, Csaszar G, Czuni L, Gurzo P, FIcsor L, Molnar B. Technical note on the validation of a semi-automated image analysis software application for estrogen and progesterone receptor detection in breast cancer. *Diagnostic Pathology*. 2011;6:6-12.

142. Aperio Technologies Inc. *Aperio Scanscope Users Guide*. 2010: MAN-0010, Revision G.

143. Brown R. Morphoproteomics: exposing protein circuitries in tumors to identify potential therapeutic targets in cancer patients. *Expert Review of Proteomics*. 2005;2(3):337-48.

144. Tan D. Morphoproteomics: a novel approach to identify potential therapeutic targets in cancer patients. *International Journal of Clinical and Experimental Pathology*. 2008;1(4):331-2.

145. Brown RE. Morphogenomics and Morphoproteomics-A Role for Anatomic Pathology in Personalized Medicine. *Archives of Pathology and133 Laboratory Medicine*. 2009;133:568-79.

146. Ferriera R, Moon B, Humphries J, Sussman A, Saltz J, Miller R, Demazo A. The Virtual Microscope. *Proceedings of the American Medical Informatics Association Annual Fall Symposia*. 1997:449-453.

147. Kayser K, Gortler J, Bogovac M, Bogovac A, Goldmann T, Vollmer E, Kayser G. AI (Artificial Intelligence) in histopathology - from image analysis to automated diagnosis. *Folia Histochemica et Cytobiologica*. 2009;47(3):355-361.

148. Rojo MG, Garcia GB, Mateos CP, Garcia JG, Vicente MC. Critical comparisonof 31 commercial available digital slides systems in pathology. *International Journal of Surgical Pathology*. 2006;14:285-305.

149. Weinstein RS, Graham AR, Richter LC, Barker GP, Krupinski EA, Lopez A m, Erps KA, Bhattacharyya AK, Yagi Y, Gilbertson JR. Overview of telepathology, virtual microscopy and whole slide imaging: prospects for the future. *Human Pathology*. 2009;40:1057-1069.

150. Jara-Lazaro AR, Thamboo TP, The M, Tan PH. Digital Pathology: exploring its applications in digital pathology practice. *Pathology*. 2010;42(6):512-518.

151. Kayser G, Razidowski D, Bzdyl P, Sommer R, Kayser K. Theory and Implimentation of an Electronic Automated Measurement System for Images Obtained from Immunohistochemically Stained Slides. *Analytical and Quantitative Cytology and Histology*. 2006;28(1):27-38.

152. Oberholzer M, Ostreicher M, Christen H, Bruhlmann M. Methods in quantitative image analysis. *Histochemistry and Cell Biology*. 1996;105:333-355.

153. Paizs M, Engelhardt JI, Siklos L. Quantitative assessment of relative changes of immunohistochemical staining by light microscopy in specified anaomical regions. *Journal of Microscopy*. 2009;234:103-112.

154. Bustin A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction. *Molecular Aspects of Medicine*. 2000;27:95-125.

155. Bustin S, Benes V, Nolan T, Pfaffl MW. Quantitative real-time RT-PCR - A perspective. *Journal of Molecular Endocrinology*. 2005;34:597-601.

156. Bustin S a, Benes V, Garson J a, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*. 2009;55(4):611-22.

157. Mullins M, Perreard L, Quackenbush JF, Gauthier N, Bayer S, Ellis M, Parker J, Perou CM, Szabo A, Bernard PS. Agreement in breast cancer classification between microarray and quantitative reverse transcription PCR from fresh-frozen and formalin-fixed, paraffin-embedded tissues. *Clinical Chemistry*. 2007;53(7):1273-9.

158. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009;27(8):1160-7.

159. Tichopad A, Kitchen R, Riedmaker I, Becker C, Stahlberg A, Kubista M. Design and optimization of reverse transcription quantitative PCR experiments. *Clinical Chemistry*. 2009;55:1816-1823.

160. Pai R, Samual P, Nehru AG, Manipadam MT, Thomas SN. Comparison of 11 endogenous control genes for normalization of mRNA obtained from paraffin-embedded tissues. *Acta Pathologica, Microbiologica et Immunologica*. 2009;117:886-892.

161. Cheung CC, Neufeld H, Lining L a, Pilavdzic D, Copete M, Garratt J, Gilks B, Torlakovic EE. The Laboratory Score/Reference Method Score Ratio (LSRSR) Is a Novel Tool for Monitoring Laboratory Performance in Immunohistochemistry Proficiency Testing of Hormone Receptors in Breast Cancer: The cIQc Experience. *American Journal of Cinical Pathology*. 2011;136(1):67-73.

162. Brooks SC, Locke ER, Soule HD. Estrogen receptor in a human cell line (MCF-7) from breast carcinoma. *Journal of Biological Chemistry*. 1973;248:6251-6253.

163. Bacus S, Kiguchi K, Chin D, King CR, Huberman E. Differentiation of cultured human breast cancer cells (AU-565 and MCF-7) associated with loss of cell durface HER-2 antigen. *Molecular Carcinogenesis*. 1990;3:350-362.

164. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo W-L, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer cell*. 2006;10(6):515-27.

165. Freshney I. *Culture of Animal Cells: A manual of basic technique and specialized applications*. Hoboken, NJ: Wiley-Blackwell; 2010.

166. Ehya H. Effusion Cytology. *Clinical Laboratory Medicine*. 1991;11(2):433-467.

167. Fischer AH, Jacobson KA, Rose J, Zeller R. Paraffin embedding tissue samples for sectioning. In: *Cold Spring Harbor Protocols*.Vol May 1.; 2008:2008 pdb.prot4987. doi: 10.1101/pdb.prot4987.

168. Kononen J, Bubendorf I, Kallioniemi A, Barlund M, P S, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallionimeni O. Tissue microarrays for high-throughput moleculat profiling of tumour specimens. *Nature Medicine*. 1998;4:844-847.

169. Dhir R. Tissue Microarrays: An Overview. In: Liu B-S, ed. *Methods in Molecular Biology Vol. 441: Tissue Proteomics: Pathways, Biomarkers and Drug Discovery*. Totana, NJ: Humana Press; 2008:91-103.

170. Jawhar NMT. Tissue Microarray: A rapidly evolving diagnostic and research tool. *Annalls of Saudi Medicine*. 2009;29(2):123-127.

171. Packeisen J, Buerger H, Krech R, Boecker W. Tissue microarrays: a new approach for quality control in immunohistochemistry. *Journal of Clinical Pathology*. 2002;55(8):613-5.

172. Sauter G. Representivity of TMA studies. *Methods in Molecular Biology*. 2010;664:27-35.

173. Garin JAL, Trull JL, Losa CA, Mas LP, Sanz LAR, Mass AG. Development of a tissue microarray (TMA) for immunohistochemical study of molecular expression profiling in prostate cancer (Part 1). *Actas Urológicas Español*. 2006;30:25-32.

174. Mirlacher M, Kasper M, Storz M, Knecht Y, Dürmüller U, Simon R, Mihatsch MJ, Sauter G. Influence of slide aging on results of translational research studies using immunohistochemistry. *Modern Pathology*. 2004;17(11):1414-1420.

175. Simon R, Mirlacher M, Sauter G. Immunohistochemical Analysis of Tissue Microarrays Simon R, ed. *Methods in Molecular Biology*. 2010;664:113-126.

176. Matkowskyj KA, Cox R, Jensen RT, Benya RV. Quantitative Immunohistochemistry by Measuring Cumulative Signal Strength Accurately Measures Receptor Number. *Journal of Histochemistry and Cytochemistry*. 2003;51(2):205-214.

177. Copete M, Garratt J, Gilks B, Pilavdzic D, Berendt R, Bigras G, Mitchell S, Lining LA, Cheung C, Torlakovic EE. Inappropriate calibration and optimisation of pan-keratin (pan-CK) and low molecular weight keratin (LMWCK) immunohistochemistry tests: Canadian Immunohistochemistry Quality Control (cIQc) experience. *Journal of Clinical Pathology*. 2011;64(3):220-5.

178. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*. 1986;14(4):1261-1295.

179. Makretsov N, Gilks B, Alaghehbandan R, Garratt J, Quenneville L, Mercer J, Palavdzic D, Torlakovic EE. Development of an evidence-based approach to external quality assurance for breast cancer hormone receptor immunohistochemistry: comparison of reference values. *Archives of Pathology and Laboratory Medicine*. 2011;135(7):874-81.

180. Mengel M, Hebel K, Kreipe H, von Wasielewski, R. Standardized on-slide control for quality assurance in the immunohistochemical assessment of therapeutic target molecules in breast cancer. *The Breast Journal*. 2006;11(1):34-40.

181. Gaffney EV. Cell and Tissue A cell line (HBL-IO0) established from human breast milk. *Cell and Tissue Research*. 1982;227:563-568.

182. Rhodes A, Jasani B, Couturier J, McKinley MJ, Morgan JM, Dodson AR, Navabi H, Miller KD, Balaton AJ. A formalin-fixed, paraffin-processed cell line standard for quality control of immunohistochemical assay of HER-2/neu expression in breast cancer. *American Journal of Clinical Pathology*. 2002;117(1):81-9.

183. Tuominen VJ, Tolonen TT, Isola J. ImmunoMembrane: a publicly available web application for digital image analysis of HER2 immunohistochemistry. *Histopathology*. 2012;60(5):758-67.

6.0 APPENDICES

6.1 Appendix A: List of chemicals and reagents used.

Ammonium hydroxide	Fisher Scientific, Ottawa, ON
Antibody dilution buffer	Dako, Goldstrup, Denmark
AU-565 cell lines	Dr. Deborah Anderson Laboratory
Autostainer IHC buffer, 10X	Dako
BT-474 cell lines	Dr. Erique Lukong Laboratory
Copper sulphate, heptahydrate	Fisher Scientific
Delimiting pen (hydrophobic marker for autostainer)	Dako
DMEM cell culture medium (Dulbecco/Vogt Modified Eagle's Minimal essential medium)	ATTC, Manassas, VA
Dimethyl sulfoxide (DMSO),	Sigma Chemical Co., St. Louis, MO
Ethylenediamintetraacetic acid (EDTA)	EMD Millipore, Billerica, MA
Envision Plus IHC stain kit for mouse IgG	Dako
Envision Plus IHC kit for rabbit IgG	Dako
Estrogen receptor antibody, rabbit monoclonal anti-human (clone SP1)	Thermo Scientific, Waltham, MA
Fetal bovine serum (FBS), cell culture grade, sterile	Gibco, Burlington, ON
Ethanol	Fisher Scientific
Formalin, 10% Buffered	Royal University Hospital (Surgipath Medical Ind., Richmond, IL)
HBL-100 cell line	Dr. Svein Carlsen Laboratory

Hematoxyllin 1	Richard Allen Scientific, Kalamazoo, MI
HemosIL Thrombin Time kit	Beckman Coulter, Mississauga, ON
HS-578T cell line	Dr. John Decoteau Laboratory
Human epidermal growth factor receptor 2, rabbit monoclonal antibody (Clone SP3)	Thermo Scientific
HUMULIN®R, biosynthetic human insulin, 100 U/mL	Eli Lilly's and Co., Toronto, ON
Human plasma, AB–	Royal University Hospital, Saskatoon, SK
Isopropanol	Fisher Scientific
αMEM cell culture medium	Gibco
nitrogen, liquid	Praxair, Saskatoon, SK
Paraplast paraffin	Surgipath Medical Ind., Richmond, IL
Isopropanol	Fisher Scientific
McCoy's 5A cell culture medium	ATTC
MCF-7 cell line	Dr. Svein Carlsen Laboratory
MDA-MB-231 cell line	Dr. John Decoteau Laboratory
MDA-MB-435 cell line	Dr. John Decoteau Laboratory
Penicillin/Streptomycin, 100X, sterile	Gibco
Permount mounting media	Fisher Scientific
pH calibration buffers (4, 7 and 10)	Fisher Scientific
Progesterone receptor antibody, mouse monoclonal anti-human (Clone 16)	Novocastra, (Leica GmbH) Concord, ON

Qiagen DNA clean-up kit*	Qiagen, Valencia, CA
Qiagen Total RNA extraction kit*	Qiagen
Roche MasterMix*	Roche Diagnostics, Indianapolis, IN
RPMI 1640 cell culture medium, 1X (Roswell Park Memorial Institute cell culture medium, 1640 variation)	Gibco
RPMI 1640 cell culture medium, 1X high glucose with 4-(2-hydroxyethyl) -1-piperazineethanesulfonic acid (HEPES) (Roswell Park Memorial Institute cell culture medium, 1640 variation)	Gibco
SKBR-3 cell line	Dr. Deborah Anderson Laboratory
Sodium bicarbonate, 100mM, sterile	Gibco
Sodium pyruvate, 100 mM, sterile	Gibco
Superscript III, reverse transcriptase*	Invitrogen, Grand Island, NY
T47D cell line	Dr. John Decoteau Laboratory
Tween 20	Dako
0.4% Trypan blue, sterile	Gibco
0.05% Trypsin/EDTA with phenol red	Gibco
Xylene, histology grade	Fisher Scientific

** Used by the Bernard Laboratory for RT-qPCR Analysis*

6.2 Appendix B: List of solutions and media used.

Bluing solution hematoxyllin enhancing solution: 0.5 mL ammonium hydroxide in 200 mL dH_20. Make fresh daily.

Cell Culture media formulations for individual cell lines: Table B1 (below) contains specific media preparations for each cell line. All media contain 100 U/mL penicillin and 100 µg/mL streptomycin.

Copper sulphate DAB enhancing solution: 2% (w/v) copper (II) sulphate in dH_20

Dako Envision Plus IHC Kit (mouse IgG kit) contains prediluted peroxidase block solution, prediluted HRP conjugated secondary anti-mouse antibody and DAB chromagen solution(1 drop DAB, 1 mL DAB buffer).

Dako Envision Plus IHC Kit (Rabbit IgG kit): contains prediluted peroxidase block solution, prediluted HRP conjugated secondary anti-rabbit antibody and DAB chromagen solution(1 drop DAB, 1 mL DAB buffer).

D-PBS (calcium and magnesium free): 4.0 g NaCl, 0.1 g KCl, 0.87 g Na_2HPO_4, dibasic, 0.1 g $KH_2PO_4 \bullet 7H_20$, monobasic dissolved in 450 mL rdH_20 water. pH solution to 7.4 (if needed), bring to 500 mL and autoclave.

D-PBS with 2 mM EDTA: D-PBS with 0.2922 g EDTA added prior to autoclaving

10 mM EDTA stock solution: 2.9224 g EDTA in 1 L dH_20.

1 mM EDTA, pH 9.0 HIER buffer: Dilute 50 mL 10 mM EDTA solution with 400 mL dH_20. pH solution to 9.0 with 1 mM NaOH and bring to 500 mL with dH_20.

70% Ethanol: 700 mL 100% ethanol and 300 mL dH_20

95% Ethanol: 950 mL 100% ethanol and 50 mL dH_20

HemosIL Thrombin Time kit: Mix 2 mL kit buffer concentrate with 4 mL sterile distilled water. Use 6 mL buffer to reconstitute 1 vial of dried thrombin. Incubate 30 min. Mix equal amounts thrombin solution with human AB plasma to create a clot.

Table B1: Media and Growth Conditions for CLMA Cell Lines

Cell Line	Growth Media	Passage
AU-565	RPMI 1640[1] high glucose with HEPES[2] + 10% FBS[3]	1:3 at <50% confluency
BT-474	DMEM[4] + 10% FBS	1:3 at 70% confluency
HBL-100	RPMI 1640 + 20% FBS	1:4-1:6 at 95% confluency
HS-578T	RPMI 1640 + 10% FBS + 0.2 U/mL Humulin®R[5]	1:4-1:6 at 95% confluency
MDA-MB-231	RPMI 1640 + 10% FBS	1:6 at 95% confluency
MDA-MB-435	RPMI 1640 + 10% FBS	1:6 at 95% confluency
MCF-7	αMEM[6] + 10% FBS + 1mM sodium pyruvate +0.01mg/L Humulin®R + 1.5 g/L sodium bicarbonate.	1:4 – 1:6 at <70% confluency
SKBR-3	McCoy's 5A[7] + 10% FBS	1:3 at <50% confluency
T47D	RPMI 1640* + 10% FBS	1:4-1:6 at 95% confluency

All media contain 100 U/mL penicillin and 100 µg/mL streptomycin.
[1]RPMI 1640, Roswell Park Memorial Institute cell culture medium, 1640 variation
[2]HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
[3]FBS, fetal bovine serum
[4]DMEM, Dulbecco/Vogt Modified Eagle's Minimal essential cell culture medium
[5]Humulin®R, Biosynthetic human insulin
[6]αMEM – α-Modified Minimum Essential Medium Eagle cell culture medium
[7]McCoy's 5A, McCoy's 5A cell culture medium

6.3 Appendix C. List of instruments and relevant materials used.

Autostain Plus IHC Autostainer System	Dako, Goldstrup, Denmark
Autostainer reagent vials, 15 mL	Dako
Biosafety Cabinet	Nuare Laboratory Equipment, Plymouth, MN
Cell culture flasks (T25, T75)	BD Falcon, Mississauga, ON
Centrifuge, benchtop	VWR Canlab, Mississauga, ON
Centrifuge tubes, 15mL and 50 mL, capped, sterile	BD Falcon
CO_2 humidified growth incubator	VWR Canlab
Coverslips	VWR Canlab
Cryopreservation chamber (Isopropanol)	Invitrogen, Burlington, ON
Cryovials	VWR Canlab
Cytospin 4	Shandon Inc., PA, USA
EZ-Slides	EMD Millipore, Billerica, MA
Filter funnels for cytospin	Biomedical Polymers Inc., Gardner, MA
Forceps	VWR Canlab
Haemocytometer	Hausser Scientific, Horsham, PA
Histological grade, tissue paper	Surgipath Medical Ind., Richmond, IL
Histology embedding cassettes	Surgipath Medical Ind.
Microarrayer, Beecher MTA-1	Beecher Instruments, Sun Prairie, WI
Microscope, reverse field	Nikon, Mississauga, ON
Microscope, regular	Nikon

Microscope slides, positively charged	VWR Canlab
Micropipettors	Eppendorf, Mississauga, ON
Microtome, Leica RM2235 -	Leica GmbH, Concord, ON
Microwave, 1100 watt	Panasonic, Mississauga, ON
Oven, incubating	VWR Canlab
Petri plates, plastic, sterile	VWR Canlab
Pipette tips, sterile, filtered	Diamed Lab Supplies Inc., Mississauga, ON
PipetteAid	Drummond Scientific Co., Broomall, PA
Pipettes, 5 mL, 10 mL& 25 mL, plastic, wrapped, sterile	Fisher Scientific, Ottawa, ON
Pipettes, glass	Fisher Scientific
Razor blades	VWR Canlab
RT-PCR Thermocycler, Lightcycler480 RT-PCR*	Roche Diagnostics, Indianapolis, IN
Sample collection containers, 125 mL, plastic, sterile	Fisher Scientific
Scanscope CS IA System	Aperio, Vista, CA
Tissue Embedding Console, Sakura	Saksura Finitech, Alphen aan den Rijn, The Netherlands
Tissue Tek TEC 5	Andwin Scientific, Schaumburg, IL
TissueTek II slide trays	Andwin Scientific
TissueTek II histology staining rack	Andwin Scientific
Tubes, 0.5 mL	Diamed Lab Supplies Inc.
Waterbath	Fisher Scientific

** Used by the Bernard Laboratory for RT-qPCR Analysis*

6.4 Appendix D: List of software used for analysis

IA Nuclear algorithm[#] (as modified by cIQc)	Aperio, Vista, CA
IA Membrane algorithm[#]	Aperio
Imagescope (Version 10)	Aperio
Microsoft Office (2007 & 2011)	Microsoft, Redmond, WA
Photoshop CS	Adobe, San Jose, CA
Profiler™ PCR Array Data Analysis Software*	Qiagen (formerly SABiosciences) Valencia, CA
Spectrum (version 11)	Aperio
SPSS (Version 19)	IBM, Armonk, NY
TMALab Version 10)	Aperio

[#]*Algorithm detailed settings are listed in Supplemental Table 1.*
**Used by the Bernard Laboratory for RT-qPCR Analysis*

6.5 Supplemental Table 1: IA Algorithms. A, Nuclear Algorithm Settings for scoring ER and PR staining. B, Membranous Algorithm settings for scoring HER2 staining. Any changes from Aperio standard algorithms are marked with bold italics.

A.

IA Nuclear Algorithm Settings	Optimized clQc Nuclear (changes in bold italics)
View Width	1000
View Height	1000
Overlap Size	100
Image Zoom	1
Markup compression type	Same as processed image
Compression quality	30
Classifier Neughborhood	0
Classifier	None
Class List	
Averaging Radius (um)	1
Curvature threshold	2.5
segmentation type	Cytoplasmic Rejection
threshold type	Edge Threshold Method
edge trimming	Weigted
lower threshold	0
upper threshold	230
minimum nuclear size (um2)	20
maximum nuclear size (um2)	1000000
mininimum roundness	0.1
minimum compactness	0
minimum elongation	0.1
remove light objects	0
Clear area intensity	240
Nuclear stain (Red)	***0.705***
Nuclear stain (Green)	***0.631***
Nuclear stain (Blue)	***0.324***
Positive stain (Red)	***0.423***
Positive stain (Green)	***0.599***
Positive stain (Blue)	***0.68***
cytoplasmic intensity threshold	230
Weak (1+) threshold	***223***
Moderate (2+) threshold	***193***
Strong (3+) Threshold	***167***
Black Threshold	0
Use Mode	Analysis/Tuning
Mark-up image type	Analysis
Classifier type	IHCNuclear
Classifier definintion module	IHCNuclearTraining

B.

IA Membrane Algorithm Settings	Optimized clQc Membrane (Same as Aperio)
View Width	1000
View Height	1000
Overlap Size	100
Image Zoom	1
Markup compression type	Same as processed image
Compression quality	30
Classifier Neighborhood	0
Classifier	None
Class List	
Averaging Radius (um)	1
Blue Curvature threshold	1
threshold type	Edge threshold method
lower blue threshold	0
upper blue threshold	220
minimum nuclear size (um2)	10
maximum nuclear size (um2)	2000
mininimum nuclear roundness	0.1
minimum nuclear compactness	0
minimum elongation	0.1
cytoplasmic correction	yes
cell/nucleus requirement	all cells
Max cell radius (um)	5
minimum cell size (um2)	30
maximum cell size (um2)	2000
minimum cell roundness	0.1
minimum cell compactness	0.1
minimum cell elongation	0.1
background threshold	240
Weak (1+) threshold	200
Moderate (2+) threshold	170
Strong (3+) Threshold	105
Completeness Threshold	50
Use Mode	Analysis/Tuning
Mark-up image type	Analysis
Classifier type	IHCMembrane
Classifier definintion module	IHCMembraneTraining

6.6 Supplemental Table 2: cIQc Run 13 Assessment results

cIQc Laboratory	ER IHC concordance with cIQc reference	ER κ[a]	PR IHC concordance with cIQc reference	PR κ[a]	PR IHC concordance with cIQc reference	HER2 κ[a]
101	35/37 (95%)	0.88	30/33 (91%)	0.82	33/36 (92%)	0.82
102	44/45 (98%)	0.95	44/45 (98%)	0.96	37/41 (90%)	0.81
103	41/43 (95%)	0.89	39/44 (89%)	0.78		
105	41/42 (98%)	0.95	38/41 (93%)	0.85	38/40 (95%)	0.9
106	47/47 (100%)	1	42/43 (98%)	0.95	36/39 (92%)	0.84
107	44/45 (98%)	0.95	40/46 (87%)	0.74	38/39 (97%)	0.95
108	38/40 (95%)	0.88	39/40 (98%)	0.95	38/38 (100%)	1
109	35/38 (92%)	0.81	36/38 (95%)	0.89	37/37 (100%)	1
110					32/38 (84%)	0.69
111	42/47 (89%)	0.74	43/45 (96%)	0.91	40/42 (95%)	0.9
112	30/32 (94%)	0.87	40/43 (93%)	0.86	40/40 (100%)	1
113	40/46 (87%)	0.67	38/39 (97%)	0.95	40/41 (98%)	0.95
114	41/42 (98%)	0.94	45/46 (98%)	0.96	42/42 (100%)	1
115	41/43 (95%)	0.89	39/42 (93%)	0.86	32/33 (97%)	0.94
116	33/40 (83%)	0.54	38/39 (97%)	0.95		
117	36/38 (95%)	0.86	40/40 (100%)	1	20/21 (95%)	0.9
118	37/41 (90%)	0.78	39/43 (91%)	0.82	35/37 (95%)	0.89
120	36/38 (95%)	0.87	28/28 (100%)	1	37/38 (97%)	0.94
122	30/43 (70%)	0.42	39/42 (93%)	0.86		
123	38/42 (90%)	0.79	32/39 (82%)	0.64		
124	35/41 (85%)	0.69	29/37 (78%)	0.58	42/43 (98%)	0.95
125	40/43 (93%)	0.84	38/43 (88%)	0.77		
126	24/29 (83%)	0.58	26/29 (90%)	0.79	28/28 (100%)	1
127	24/26 (92%)	0.82	23/24 (96%)	0.91	31/31 (100%)	1
128	28/29 (97%)	0.92	33/34 (97%)	0.94		
129	32/35 (91%)	0.82	30/34 (88%)	0.77	33/34 (97%)	0.94
132	24/27 (89%)	0.78	28/30 (93%)	0.87		
133	31/34 (91%)	0.81	35/36 (97%)	0.94	35/35 (100%)	1
134	21/35 (60%)	0.14	19/37 (51%)	0.03	35/37 (95%)	0.89
135	40/43 (93%)	0.85	41/42 (98%)	0.95	38/41 (93%)	0.85
138	33/36 (92%)	0.82	34/39 (87%)	0.75		
139	36/38 (95%)	0.88	34/35 (97%)	0.94	25/25 (100%)	1
140	32/35 (91%)	0.82	35/36 (97%)	0.94	37/39 (95%)	0.89
141	38/43 (88%)	0.74	40/44 (91%)	0.81		
143	39/41 (95%)	0.89	32/35 (91%)	0.83		
144	41/44 (93%)	0.85				
145	36/41 (88%)	0.72	38/40 (95%)	0.9	40/40 (100%)	1
146	35/37 (95%)	0.89	35/36 (97%)	0.94		
147	42/48 (88%)	0.73	35/37 (95%)	0.89	29/36 (81%)	0.62
148	43/44 (98%)	0.95				
149	22/24 (92%)	0.83	41/45 (91%)	0.82	40/43 (93%)	0.86
150	41/44 (93%)	0.84	37/38 (97%)	0.95	34/34 (100%)	1
151	30/39 (77%)	0.47	31/43 (72%)	0.45	36/40 (90%)	0.79
153	28/33 (85%)	0.68	38/42 (90%)	0.81	37/39 (95%)	0.89
154	33/36 (92%)	0.82	35/36 (97%)	0.94		
155	34/37 (92%)	0.82	32/33 (97%)	0.94	32/33 (97%)	0.94
156	33/35 (94%)	0.88	34/36 (94%)	0.89	36/37 (97%)	0.94
157	40/40 (100%)	1	37/42 (88%)	0.76	35/36 (97%)	0.94
158	39/41 (95%)	0.89	41/42 (98%)	0.95	38/40 (95%)	0.9
159	33/35 (94%)	0.88	37/39 (95%)	0.9		
160	34/35 (97%)	0.94	32/35 (91%)	0.83	31/32 (97%)	0.93
161	25/27 (93%)	0.83	26/27 (96%)	0.92	28/28 (100%)	1
164	27/32 (84%)	0.66	30/30 (100%)	1	31/31 (100%)	1
167	19/21 (90%)	0.8	24/28 (86%)	0.72	24/24 (100%)	1
168	21/23 (91%)	0.81	21/22 (95%)	0.9		
170	35/38 (92%)	0.81	25/27 (93%)	0.85	33/34 (97%)	0.94
172	23/31 (74%)	0.42	28/30 (93%)	0.86		
173	26/30 (87%)	0.71	30/34 (88%)	0.76		
174	29/32 (91%)	0.81	28/32 (88%)	0.75		
175	30/32 (94%)	0.87	30/32 (94%)	0.88	27/28 (96%)	0.92
177	25/27 (93%)	0.83	25/30 (83%)	0.67		
178	30/34 (88%)	0.75	27/34 (79%)	0.6		
% Acceptable	66%	69%	75%	73%	95%	93%

[a]Cohen's κ calculated between laboratory IHC results and cIQc IHC assessed staining results